Deanship of Graduate Studies

Al-Quds University

Hybrid Artificial Intelligence Model for Prediction and Classification Across all Stages of Brain Ischemic Stroke in Non-enhanced Computerized Tomography Images

Student: Ibraheem Bassam Ibraheem Qdaih

M.Sc. Thesis

Jerusalem- Palestine

1445/2024

Hybrid Artificial Intelligence Model for Prediction and Classification Across all Stages of Brain Ischemic Stroke in Non-enhanced Computerized Tomography Images

Student:
Ibraheem Bassam Ibraheem Qdaih

B.Sc. of Medical Imaging / An-Najah National University / Palestine

Supervisor: Dr. Radwan Qasrawi

This Thesis submitted in Partial Fulfilment of Requirements for the Degree of Master of Medical Imaging Technology Track Functional Imaging/ Faculty of Health Professions/ Al-Quds University

Jerusalem- Palestine

1445 / 2024

Al-Quds University

Deanship of Graduate Studies

Faculty of Health Profession

Medical Imaging Technology

Thesis Approval

Hybrid Artificial Intelligence Model for Prediction and Classification Across all Stages of Brain Ischemic Stroke in Non-enhanced Computerized Tomography Images

Prepared by: **Ibraheem Bassam Qdaih**

Registration No: **22010787**

Supervisor: **Dr. Radwan Qasrawi**

Master Thesis submitted and accepted, Date: 26/ 05/ 2024

The names and signatures of the examining committee members are as follows:

1- Head of Committee: **Dr. Radwan Qasrawi** Signature:

2- Internal Examiner: **Dr. Hussein ALMasri** Signature:

3- External Examiner: **Dr. Iyad Tumar** Signature:

Jerusalem – Palestine

1445 / 2023

Dedication

This research is dedicated:

To Mother, Father, and lovely Brothers.

To the absent friend soul, my uncle Hussam Qdaih.

To every Martyr and Prisoner in the bereaved land of Palestine.

To my Motherland, Gaza, and Jerusalem.

To my path companions, Omar and Suhaib

To my Family, Instructors, Colleagues, and Friends.

Declaration:

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or Institution.

Signed: 

Ibraheem Bassam Ibraheem Qdaih

Date: 26/ 05/ 2024

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful, at the beginning and the end. And the Peace and Blessings be upon our Prophet Mohammad Al-Hadi Al-Amin.

First and foremost, my deepest gratitude to Allah for everything in my life, for his guidance which transcends all understandings, in composing this thesis. Without His unwavering support, I would not have been able to accomplish what I have today. I am profoundly grateful to the source of motivation, my dear parents, and my lovely brothers, for their endless love, patience, encouragement, and sacrifices throughout this endeavor. Their unwavering belief in me has been my source of strength. I greatly appreciate my esteemed supervisor, Dr. Radwan Al-Qasrawi, for his invaluable guidance, mentorship, and support. His insightful feedback and encouragement have been instrumental in shaping this thesis.

To those who shared burdens, from the darkest of nights to the lightest of days, where every step is laden with challenges. Omar and Suhaib, together we've weathered storms, overcome obstacles, and celebrated successes, no hurdle is insurmountable with you by my side.

Lastly, Thanks to all physicians for participating and cooperating in making this study and to all those who have facilitated my path and contributed to the completion of this thesis in any way. Your assistance and encouragement have been deeply appreciated.

Hybrid Artificial Intelligence Model for Prediction and Classification Across all Stages of Brain Ischemic Stroke in Non-enhanced Computerized Tomography Images

Student: Ibraheem Bassam Ibraheem Qdaih

Supervisor: Dr. Radwan Al-Qasrawi

Abstract

Stroke is a major global health issue, resulting in significant mortality and disability among approximately 16 million people annually. Rapid response is crucial to mitigate brain damage and improve patient outcomes. Strokes, which are primarily categorized into ischemic and hemorrhagic types, vary in presentation and can be influenced by modifiable risk factors. In regions like Palestine, with limited economic resources, stroke is a prevalent cause of death. Diagnostic challenges are heightened by the limitations of brain non-enhanced CT (B-NECT) scans, which vary in effectiveness based on the stroke's stage.

This study introduces a novel artificial intelligence-based framework, the Stroke Precision Enhancement Model (SPEM), which employs image processing, deep learning, and machine learning techniques to enhance the classification of ischemic stroke stages in B-NECT images. This real-time hybrid model integrates Contrast Limited Adaptive Histogram Equalization (CLAHE) for preprocessing, with feature extraction conducted through Densely Connected Convolutional Networks-121 (DenseNet-121). Classification is performed using Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR), with a focus on determining the most effective method based on various performance metrics.

The results indicate exceptional performance of the SPEM, especially when combining DenseNet-121 with the LR classifier. Notably, in the hyper-acute stage, the model achieved an accuracy of 0.9957, a precision of 0.9914, and a remarkable Area Under the Receiver Operating Characteristic Curve (AUC) of 0.9999, with a processing time of just 0.04 seconds. Similar high performance was maintained across other stroke stages. These findings highlight the potential of this AI-enhanced model in facilitating faster and more accurate clinical decisions for early-stage stroke treatment.

The hybrid model shows promise in predicting and classifying ischemic strokes and could significantly impact clinical practice upon further research, validation on larger datasets, enhanced interpretability, and integration into clinical workflows.

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations and Units

	Page
A	
(AIS) Acute Ischemic Stroke	2
(AI) Artificial Intelligence	6
(ASPECTS) Alberta Stroke Program Early CT Score	8
(AUC) Area Under the ROC Curve	8
(ABTD) Analysis of Brain Tissue Density	15
(ANN) Artificial Neural Network	17
(ADC-MRI) Apparent Diffusion Coefficient Magnetic Resonance Imaging	30
B	
(B-NECT) Brain Non-enhanced CT	2
(BR) Border Regions	35
C	
(CVA) Cerebrovascular Accident	1
[CTs] Computed Tomography scans	2
(CAD) Computer-Aided Diagnostic	6
(CNN) Convolutional Neural Network	6
(CLAHE) Contrast Limited Adaptive Histogram Equalization	8
(CLAHE-4) Contrast Limited Adaptive Histogram Equalization at Clip Limit 4	53
(CTP) CT Perfusion	12
(CTA) CT-Angiography	13
(CBF) Cerebral Blood Flow	13
(CBV) Cerebral Blood Volume	13
(CAROI) Circular Adaptive Region of Interest	15
(CM) Central Moment	15
[CI] Confidence Interval	17
(CSF) Cerebrospinal Fluid	24
(CR) Corner Regions	35

Chapter One : Introduction

1.1 Background and Context

Being one of the leading causes of death and disability worldwide, stroke disease affects approximately 16 million persons annually around the world (Johnson (2016). Stroke, or cerebrovascular accident (CVA) was called 'apoplexy' before 2400 years (JE., 1996), is a neurological disease that can lead to motor and cognitive impairment or death due to blood perfusion lack in some area of the brain. Oxygen absence within few minutes in specific area lead to irreversible neuronal death called "core" or infraction, the surrounding area are non-functioning called "penumbra" that retains only the blood flow from alternative blood flow pathway to avoid cells death (Mozaffarian D, 2016). So, effective treatment to minimize the penumbral areas is very important and needs rapid diagnosis and management before death by reperfusion therapy (Saver, 2017).

Strokes can be classified into ischemic strokes or haemorrhagic strokes; ischemic strokes occur when blood vessels are occluded by a thrombus or embolus that constitutes 70% of all strokes, while haemorrhagic strokes (12%) are caused by rupture and bleeding of a blood vessel in the surrounding brain tissue (Johnson, 2016). Several risk factors, both modifiable and non-modifiable, have been identified in stroke disease, that are attributable to age, gender, ethnicity, history of cardiac disease, sedentarism, obesity, poor nutrition, tobacco use, and alcohol consumption. Modifiable risk factors may be identified, controlled, and treated in individuals for the prevention and detection of stroke disease.

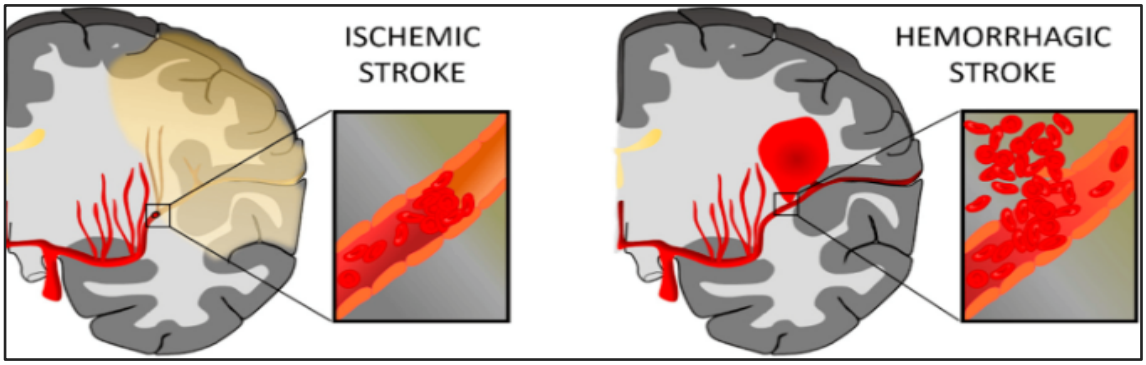


Figure 1: Comparative Representation of Ischemic and Hemorrhagic Strokes. Figure created with BioRender.com.

In Palestine which is considered a low and middle income-country (LMIC) or peripheral, lower-developed country, with a population of 5,324,656 inhabitants, as reported by the Ministry of Health (MOH), in 2022 stroke was considered the fourth death cause (10.5%) after Diabetes Mellitus (DM) in West Bank (WB), the third causes of death (11.6%) in Gaza Strip (GS) after Malignant Neoplasm (MOH, 2023).

Categorizing the type of stroke is very important task, due to the radical difference in occurrence mechanism and treatment. Stroke diagnosis is based on a detailed medical history, a physical and neurological examination, and brain imaging testing (Computed Tomography scans [CTs] or Magnetic Resonance Imaging [MRIs]). Stroke classification based on obtained imaging and patient data is essential to determine treatment and rule out other stroke mimics, such as brain tumours, and subdural hematomas, among others. Thus, accurate and early detection of stroke, as well as prediction of its long-term outcomes, is key for treatment selection and managing prognostic expectations (Heo, 2019), especially at the time window of 4.5 hours which considered the optimal window to obtain improved treatment outcomes (Emberson J, 2014).

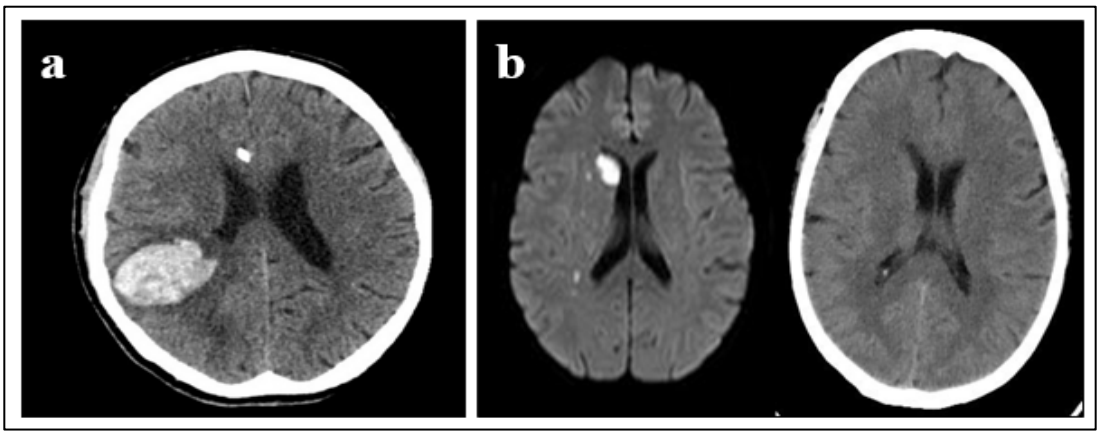


Figure 2 (a, b): Comparison of Hemorrhagic and Ischemic Stroke (a) B-NECT image showing evidence of hemorrhagic stroke. (b) Comparison of DWI-MRI (left side) and B-NECT (right side) in detecting hyper-acute ischemic stroke.

The ultimate therapeutic objective in the management of ischemic stroke and brain infarcts is to minimize neuronal damage by alleviating arterial blockages through recanalization and restoring healthy cerebral blood flow via reperfusion (Prabhakaran et al., 2015). The fundamental hypothesis underpinning the pathophysiological response during acute ischemic stroke (AIS) treatment posits that once a cerebral artery becomes obstructed, brain tissues with reduced blood flow are susceptible to permanent infarction. However, these at-risk tissues can

be effectively salvaged by swiftly restoring normal blood flow (Zhao et al., 2013). These vulnerable regions are referred to as ischemic penumbras, and the primary aim of AIS treatment is to prevent their progression to irreversible infarction (Prabhakaran et al., 2015). The primary challenges in stroke treatment stem from the need to accurately distinguish the penumbra from the core zone and differentiate the penumbra from benign hypo-perfused tissues, which experience poor perfusion without the risk of infarction (Kidwell et al., 2003).

Compared with all available medical imaging modalities, CT is considered the widest, fastest, and most prevalent modality to exclude brain haemorrhages (Jovin TG, 2015). The ischemic stroke appearance is different based on the time in which the patient is scanned relative to symptoms onset [hyper-acute (less than 4.5 hours), acute (less than 24 hours), sub-acute (24 hours to 7 days), chronic (weeks)](Dzialowski I, 2007). The hypoattenuation appearance of the ischemic region in the brain means irreversible (died) infracted tissue (Dzialowski I, 2007). In the hyper-acute stage, specific relevant features indicate the evidence of CVA as hyperdense middle cerebral artery (MCA), inability to differentiate cortical grey-white matter interface (N. Tomura, 1988). However, due to the decreased brain tissue attenuation, the poor contrast, and signal-to-noise ratio in CT, it's difficult to detect the AIS area (D. A. Barber PA, Zhang J, Buchan AM, 2000).

Hounssfield Unite (HU), is a relative quantitative measurement of the tissue density. By the difference of the tissue density, different X-ray beam energies will produced due to the attenuation, and or absorption events that result from photon interaction with the body tissues. With water having a value of zero HU, any tissue denser than water will have a positive HU value, and any tissue less dense than water will have a negative value. For brain tissue, the HU value ranged from +20 to +40. As reported by The Radiology Assistant, each 1% of water content in the brain tissue will decrease the 2.5 of HU. Ischemic stroke is represented as hypoattenuation (the HU value decreases from the normal or relatively becomes closer to zero), due to the increased of water content as a result from the cytotoxic edema developed from the died cells (Thurnher, 2008).

In the hyper-acute stage (within 6 hours of onset), CT may appear normal in the first few hours as signs such as loss of Gray-White differentiation have not developed (Schaefer et al., 2003). Meanwhile, DW-MRI is highly sensitive for delineating the ischemic core region due

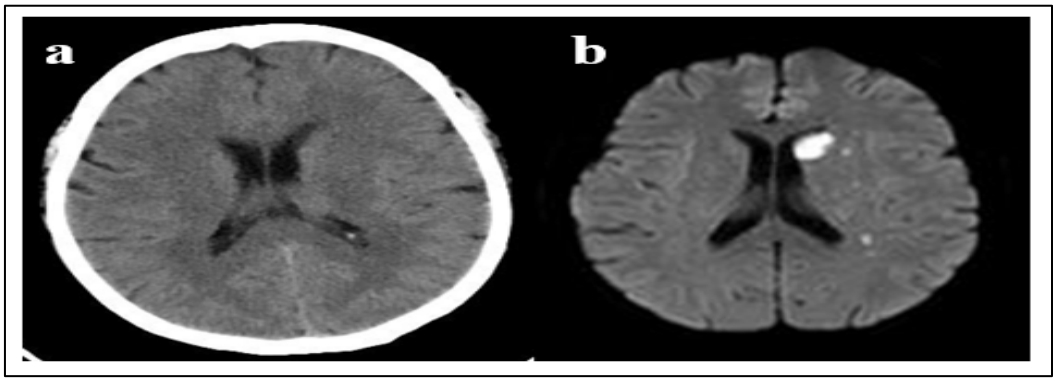


Figure 3 (a, b): Normal B-NECT showed (a). However, DW-MRI (b) highlights a left caudate nucleus restricted region, indicative of hyper-acute ischemic stroke (four hours after symptom onset).

to restricted diffusion appearing as hyperintensity on sequences, allowing accurate demarcation of the infarcted tissue (Warach et al., 1992).

As time progresses, CT sensitivity increases such that in the acute stage (6-72 hours), hypoattenuation becomes visible on CT scans (Barber et al., 2000). This allows for 80-90% detectability of infarcts between 1-2 days after stroke onset. During this same acute period, DW-MRI with perfusion MRI (pMRI) can identify the ischemic penumbra through a mismatch between the irreversible ischemic core seen on DWI and the potentially salvageable ischemic tissue within the perfusion lesion on pMRI sequences (Parsons et al., 2012).

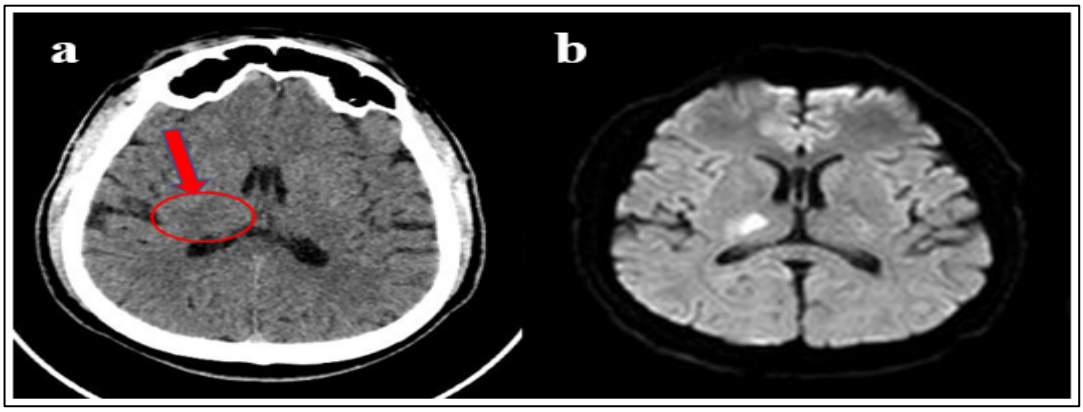


Figure 4 (a, b): B-NECT (a) reveals minimal density change in the right lentiform nucleus (red cercle). (b) DW-MRI highlights the right lentiform nucleus restricted region indicating acute ischemic stroke lesion (20 hours after symptom onset).

As the stroke continues into the sub-acute stage (3-10 days), CT hypodensity progresses and edema peaks at around 5-7 days (Barber et al., 2000). Concurrently, T2-Fluid-attenuated inversion recovery (T2-FLAIR MRI) sequences demonstrate hyperintensity developing in the infarcted region (Warach et al., 1992). Figure (5, a) B-NECT showed left frontal lobe hypodensity edema became evident after 5 days from the onset of the symptoms, indicative of sub-acute ischemic stroke. While in figure (5, b) represents a T2-FLAIR MRI image a week from the symptom onset, showing cortico-subcortical infraction in the right middle cerebral artery territory, the image indicates perifocal edema leading to a compression of the adjacent lateral ventricle.

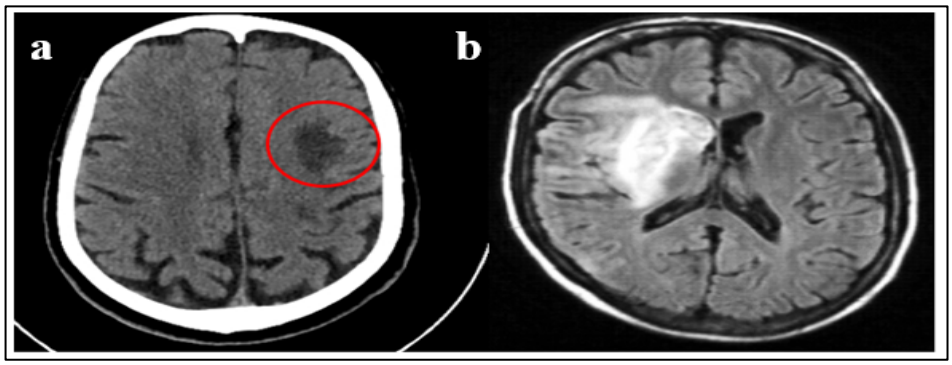


Figure 5 (a, b): (a) B-NECT and (b) T2-FLAIR MRI (Ritzl et al., 2004), sub-acute ischemic stroke lesions.

Finally, in the chronic stage (>3 months), CT shows atrophy and encephalomalacia as hypodensity evolves (Barber et al., 2000). Meanwhile, volume loss occur over months on sequences to depict the mature infarct (Warach et al., 1992).

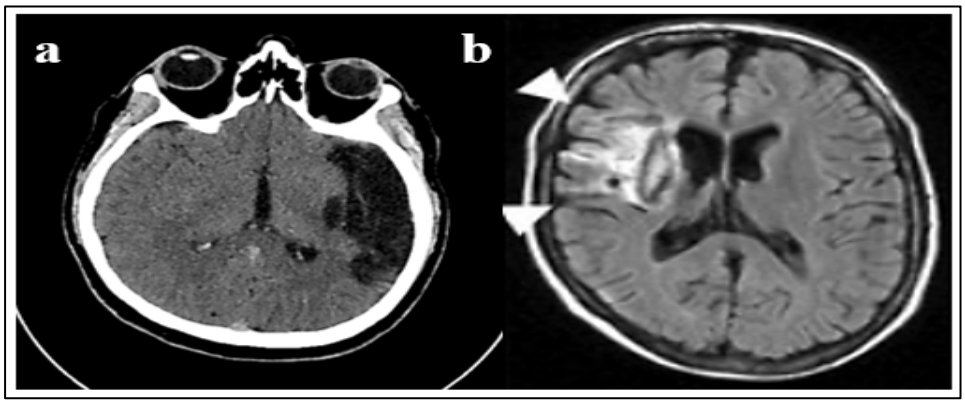


Figure 6 (a, b): (a) B-NECT showed hypodensity chronic ischemic stroke in the left MCA territory indicating (4 months), (b) T2-FLAIR MRI (Ritzl et al., 2004), right MCA territory chronic ischemic lesion.

Concerning stroke, with the development of computer capabilities, computer-aided diagnostic (CAD) systems have been utilized to assist physicians in stroke prevention, prediction, diagnosis, and treatment planning in a short time, and so save the urgent time needed for brain tissue survival (Tyan et al., 2014).

Artificial Intelligence (AI) has recently inundated the field of healthcare as electronic data accessibility and technological advances expand. AI can make use of patient data, including that stored in the electronic health record (EHR) databases, clinical, imaging, and laboratory datasets to predict disease incidence and prognosis(Cheon, 2019). Machine learning (ML) algorithms, in particular, have been studied and yielded exciting results for the classification and prediction of chronic diseases. Furthermore, ML can also be used in interpreting medical imaging as it recognizes patterns of imaging information and renders medical diagnosis (Lee, 2017).

Recently, deep learning (DL) considered the gold standard and the most widely used computational approach in the ML field, which simulates the human brain process to extract distinctive features (LeCun Y, 2015). Convolutional neural network (CNN) is considered the most common DL used segmentation and detection algorithm in medical images because it identifies features automatically without the need for human supervision, used in computer vision tasks (Russakovsky O, 2015). CNN consists of multi-layers, the first layer is the convolutional layer that performs feature extraction, the second layer is a pooling layer for the sub-sampling of the feature map and the third layer is the fully connected layer (FC) located at the end of CNN convert the input from the convolutional and pooling layers to vectors of one dimension (Gu J, 2018).

The use of machine learning and medical image processing in the detection and prediction of stroke disease can enhance precision medicine, individualized treatment, and timely detection and accurate classification of the disease. Thus, decreasing the risk of stroke, increasing survival rates as well as improving the quality of life of stroke patients

1.2 Problem Statement

Ischemic stroke is due to occlusive process of a specific cerebral artery leads to lack of blood perfusion to the brain tissue that the artery supplies it which causes early-stage symptom as Amnesia (i.e., memory loss including facts, knowledges and experiences), Dementia (i.e., inability to think or make decisions), extremities weakness or numbness, visual or hearing loss,

and other symptoms. So, treatment is urgent and any delay in the treatment process will cause irreversible damage to the penumbra (i.e., the tissue under the risk). In general, fast treatment needs fast diagnosis, commonly the diagnosis starts with B-NECT and/or brain MRI, however, MRI is higher cost and longer time compared with CT. Moreover, in LMIC regions MRI is not available in all clinical centers and hospitals - in WB, for example, only 3 MRI services are available in the governmental hospitals - (PMOH, 2018). Despite the ionizing radiation and the limited ability to CT to detect early ischemic changes, CT is an abundant and very fast technique and lower cost. The limited ability to detect AIS accurately in B-NECT is due to the poor contrast, and signal-to-noise ratio (SNR) of CT.

As reported by Khatib et al.2018, 150 patients with acute ischemic stroke patients admitted to two public hospitals from September 2017 to May 2018 in WB of Palestine, had B-NECT scans, and 98% of them received antiplatelet therapy to prevent recurrent stroke events without thrombolysis or thrombectomy treatment due to unavailability to localize the ischemic region accurately, 23% of patients have at least one post-stroke complication (Khatib et al., 2018).

Since the past few years, there have been dramatic changes in acute stroke treatment (Menon BK, 2009). In acute stroke, the ultimate New mechanical thrombectomy devices were introduced for recanalization of the occluded vessels (H. M. Menon BK, Eesa M, et al., 2011). Even with this these advanced techniques, many stroked patients undergo treatment do not do well clinically (G. M. Menon BK, 2011). Nonetheless, many studies showed that improved clinical outcomes are reached if the brain recanalized as fast as possible when the brain tissues are reversible (Rha JH, 2007). Menon et al. in 2014 found in their study that the rate of bad outcomes increased by 14% for every 30-minute delay in treatment (Menon BK, 2014). Thus, an ideal technique enables to detect of acute ischemic changes in the brain quickly and reliably should be available.

Based on the above-mentioned factors, a discriminative framework was built to predict, detect stroke evidence utilizing image pre-processing algorithm with DL approach combined with ML classifier to predict and classify strokes in B-NECT images in all CVA stages (hyper-acute, acute, subacute and chronic), which provide fast and accurate diagnosis enable the urgent and rapid treatment plan.

1.3 Study Significance

Multidisciplinary efforts have been advanced over the last two decades, to improve acute stroke diagnosis and management (Lin CB, 2012) by developing diagnostic tools such as The National Institute of Health Stroke Score (NIHSS) (Josephson SA, 2006; Lyden P, 2009), the modified Rankin scale (mRS) (Sharath KG, 2018), and Alberta Stroke Program Early CT Score (ASPECTS) (Puetz V, 2008; Tei H, 2010). Moreover, new thrombectomy devices with a multidisciplinary team to optimize the workflow (Saver JL, 2016).

Several studies have been focused on using AI to improve the prediction and detection of acute ischemic stroke on B-NECT (Chiang et al., 2022; Finck et al., 2022; Hanning, 2020; Kniep et al., 2020; Lin et al., 2022; Lombardi et al., 2018; Peter et al., 2017; Reboucas Filho et al., 2017; Wu Qiu, 2020). However, the significance of this work lies in its comprehensive approach to improving the diagnostic accuracy for ischemic stroke. By integrating contrast limited adaptive histogram equalization (CLAHE) as a pre-processing technique on B-NECT and Densely connected convolutional networks-121 (DenseNet-121) as a feature extraction DL technique and ML classifiers for classification and prediction task, this research offers a novel framework for the precise classification of ischemic stroke into its various stages: normal, hyper-acute, acute, sub-acute, and chronic. This is particularly groundbreaking as previous studies have not focused on the prediction and classification of ischemic stroke across all these stages.

In addition, this is the first study in Palestine, and to our knowledge, very limited studies globally assess the impact of using DenseNet-121 feature extraction in the original dataset without and with CLAHE pre-processing. By this two-evaluation fold, good understand can be achieved precise the role of image enhancement in the model performance. Moreover, the study validating the effectiveness of various machine learning classifiers to determine the optimal classifier for real-time prediction and classification.

Crucially, the performance of the proposed model is evaluated using a robust set of metrics including accuracy, precision, recall, f1-score, and area under the receiver operating characteristic (ROC) curve (AUC), providing a thorough understanding of its effectiveness. This multifaceted evaluation is unprecedented in the existing literature. Furthermore, by considering elapsed time as a metric, this research spearheads the drive towards real-time models in medical diagnostics. The inclusion of time efficiency as a key performance metric underlines the aim to develop a practical, deployable system that can operate within clinical

time constraints, thus enhancing its potential for real-world application. Altogether, the significance of this study is profound, as it not only furthers the scientific community's understanding of effective machine learning applications in medical imaging analysis but also paves the way for more timely and accurate stroke diagnosis applications, with potential region-wide and global implications for patient care and outcomes.

1.4 Study Goal

The primary goal of this study is to enhance the accuracy and timeliness of ischemic stroke diagnosis through the implementation of an advanced artificial intelligence-based framework, specifically tailored to improve the classification of stroke stages using brain non-enhanced CT (B-NECT) images.

1.5 Study Objectives

Objective 1: Develop and Implement the Stroke Precision Enhancement Model (SPEM).

- Integrate Contrast Limited Adaptive Histogram Equalization (CLAHE) for preprocessing and Densely Connected Convolutional Networks-121 (DenseNet-121) for feature extraction to create the SPEM.
- Employ machine learning classifiers such as Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) within the SPEM to evaluate their effectiveness in stroke classification.

Objective 2: Assess the Performance of SPEM.

- Evaluate the accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC) of SPEM in classifying different stages of ischemic stroke.
- Compare the performance metrics of SPEM with traditional diagnostic methods to determine improvements in diagnostic capabilities.

Objective 3: Analyze the Real-time Application Feasibility of SPEM.

- Measure the processing time of the SPEM under different conditions and stages of stroke to assess its suitability for real-time clinical applications.

- Identify any potential operational challenges in deploying SPEM in a clinical setting and propose solutions to optimize its integration.

Objective 4: Validate the Clinical Efficacy of SPEM in Diverse Settings.

- Conduct a pilot study in a healthcare setting with limited resources, such as in Palestine, to test the adaptability and effectiveness of SPEM across different economic environments.

1.6 Research Hypothesis

Through this work, we hypothesized that we will achieve the following aims:

- The Stroke Precision Enhancement Model (SPEM) that employs Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing and DenseNet-121 for feature extraction will perform significantly better in classifying the stages of ischemic stroke in B-NECT images than traditional image processing techniques without AI enhancement.
- The SPEM, when utilizing ML models for classification, will achieve higher accuracy, precision, and area under the receiver operating characteristic curve (AUC) in the hyper-acute stage of ischemic stroke compared to using traditional models.
- The use of the SPEM will reduce the processing time for classifying stroke stages in B-NECT images, making it feasible for real-time application in clinical settings, thereby facilitating faster and more accurate clinical decisions for early-stage stroke treatment.
- The diagnostic accuracy of SPEM in the classification of ischemic stroke stages is not significantly compromised by the economic limitations of healthcare settings in regions like Palestine, demonstrating the model's robustness and adaptability across different healthcare environments.

Chapter Two: Literature Review

This chapter includes a review of several published scientific researches on stroke management and diagnosis, and different methods including neuroimaging and AI techniques used to improve ischemic stroke patients interpolation on different medical imaging modalities.

2.1 Brief Perspective on Stroke Neuroimaging Studies

Neuroimaging plays a pivotal role in determining the best way for patient who suffering from acute ischemic stroke. The importance of imaging is to rule out brain haemorrhages and stroke mimics as tumours, localization of ischemic core and penumbra presence and embolus identification.

B-NECT is considered the primary line of neuroimaging tools due to the wide availability of CT. Moreover, CT is considered a cost-effective modality used to diagnose large major strokes, reversible and irreversible brain tissue, stroke mimics, and haemorrhages. However, the sensitivity and specificity of the B-NECT scan to detect ischemic changes is 60%, 85% retrospectively within the first 6 hours after symptoms onset, while the sensitivity increased to 100% after 24 hours from symptom onset (Barber PA, 2005).

In 1999 Michal Lev and his research partners used variable window width (WW) and window centre (WC) to improve CT sensitivity and specificity to detect AIS of 21 patients suffered from acute stroke (<6 hours), they compared wide WW of 80 Hounsfield unit (HU), WC of 20 HU, and narrow window (WW: 8HU, WC: 8 HU). The results showed decreased sensitivity (57%) when using wide window while the sensitivity increased to (71%) when the narrow window. The specificity in both window is 100%. P value = 0.03. (Lev et al., 1999).

Shraddha Mainali 2014, used narrow window width named it "Stroke window" to detect AIS of 50 B-NECT at setting of WW 30 HU and WC of 35 HU, stroke window results were compared with the standard brain window (WW: 100 HU, WC: 35 HU). In this study 35 early ischemic changes were detected by stroke window, while 9 early ischemic changes were detected by standard window (70% Vs. 18%, P<0.0001). Stroke window improve the early ischemic lesion detection (Shraddha Mainali, 2014).

Recent advances in neuroimaging have raised the hopes for accurate detection of early ischemic changes with better sensitivity, particularly DW-MRI. Srinivasan et al. 2006, Dr

Barber et al. 2005, and Smajlović and Sinanović 2004 studied the sensitivity of DW-MRI and B-NECT of acute stroke patients and the result showed higher sensitivity of DW-MRI (98.6% Vs. 60%). In contrast, CT is faster and the most available neuroimaging modality for acute stroke detection. Moreover, MRI is susceptible to motion artifacts and is time-consuming. (Barber PA, 2005; Nael et al., 2014; Smajlović & Sinanović, 2004; Srinivasan et al., 2006).

It has, however, been reported that brain CT Perfusion (CTP) which is a scanning protocol enables to evaluate of the perfusion within brain tissue, by providing a dynamic evaluation of brain capillary wash-in, and wash-out of the injected iodinated contrast media intravenously, thus increasing the confidence to diagnose ischemic stroke (Donahue & Wintermark, 2015; Konstas AA, 2009). It was published in 2019 by Becks et al. that the accuracy of CTP to detect intracranial vessel occlusion is significantly comparable (AUC; 0.92-0.97) with sensitivity ranging between 91%-96% and specificity of 86%-96% (Becks et al., 2019).

ASPECT is a quantitative score innovated by Barber et al. in 2000 to evaluate the extent of hyper-acute ischemic changes by dividing the anterior and middle cerebral artery regions into 10 regions each sign has a point value (D. A. Barber PA, Zhang J, Buchan AM., 2000). American Heart Association recognized ASPECT for acute stroke management and recommended that the baseline of ASPECT ≥ 6 needs endovascular therapy (Yoo et al., 2016).

Studies reported that ASPECT usage advantage is that advanced imaging modalities such as CTP and pMRI are not needed. Despite the difference in agreement between the physicians depending on the reader's experience, work station quality which can alter the interpretation of ASPECT, and so, the decision accuracy (Mainali et al., 2014; Wardlaw et al., 2007)

Menon et al. in 2015, developed a new imaging tool faster than CTP enables the clinicians to have more knowledge about the degree of occlusion by detecting the degree of contrast filling the artery all of brain region relative to time manner. Multiphase CT angiogram (mCTA) that improve the detectability of occlusive process by observing "the delayed vessel sign" (Byrne et al., 2017). this sign refers to the delay of vessel enhancement located distal to the occlusion site. This technique provides excellent inter-rater reliability ($n = 30$, $k = 0.81$, $P < 0.001$) and AUC of 0.72 which is better than CTA (0.64) as reported by Wang et al., 2020 (Menon et al., 2015; Z. Wang et al., 2020).

Despite B-NECT is widely available, cost-effective, and a very fast technique, B-NECT interpretation depends on the expert's skill. Moreover, B-NECT is considered insensitive to detect small stroke lesions (<20%) in the first 3 hours from symptom onset due to the low spatial resolution (Musuka et al., 2015; Rudkin et al., 2018). While CT-angiography (CTA) is an available technique in many hospitals, it can provide accurate detection when the occlusion is related to large vessels, but it's related to physician experts when the occlusion in small vessels, also it needs time for interpretation. Moreover, it is recommended that CTA should be immediately followed by B-NECT especially when the occlusion is suspected in the vertebrobasilar artery (Musuka et al., 2015; Rudkin et al., 2018; Vagal et al., 2016). As for the mCTA technique, it gives accurate detection of the distal brain arteries with better inter-rater reliability, despite the high experience needed for interpolation and the delaying time for image acquisition (Menon et al., 2015; Volny et al., 2017).

CTP is comparable to CTA and mCTA, because CTP provides color maps for the parenchymal enhancement within time, in other words, we can say that CTP assesses the tissue of the brain at the capillary level, by measuring the cerebral blood flow (CBF), cerebral blood volume (CBV), time to peak (TTP) enhancement and time to maximum (Tmax) enhancement, thus means that no need to high experience for interpretation, but, too much time is needed for image acquisition (40-90 seconds) of continuous scanning which make it vulnerable to motion artifact (Campbell et al., 2012; Ospel et al., 2020).

We can't ignore that CTA, mCTA, and CTP use higher radiation doses, The effective dose of mCTA was 5.73 mSv, which was equal to that in PCT, and it was 3.57 mSv in CTA (Yang et al., 2008). Also, iodinated contrast media injection is part, which means that not all patients fit for these techniques especially when the patient has cognitive renal dysfunction, heart diseases, DM, anemia and/or renal transplant, or has a history of allergic reactions from iodine which extend from mild to life-threatening emergency (Lightfoot et al., 2009).

2.2 Artificial Intelligence (AI) Previous studies

2.2.1 Computer-Aided Diagnosis

In the field of medical imaging, CAD systems have emerged as valuable tools for assisting healthcare professionals in the detection and interpretation of abnormalities. One area where CAD has shown significant promise is in the detection of ischemic stroke, a critical medical condition that requires timely diagnosis and intervention to minimize brain damage

and improve patient outcomes. The integration of mathematical models into CAD schemes has shown promise in improving the accuracy of disease detection by identifying abnormalities that clinicians may overlook. This approach has achieved notable success in the field of radiological science, due to several key components which include image pre-processing techniques to enhance image quality, feature extraction methods to capture relevant information, and machine learning algorithms to classify and predict the presence of ischemic stroke, so that, it's possible to quantify and combine the tiny pixel-wise correlation between tissue (Berkhemer et al., 2015; Liao et al., 2006; Maroulis et al., 2005; Reboucas Filho et al., 2017; Vogelsang et al., 1998).

CAD systems in medical imaging utilize advanced algorithms and techniques extend from machine learning, image processing, and pattern recognition to analyse medical images and identify potential abnormalities to more advanced DL models as CNN and recurrent neural network (RNN). In the context of ischemic stroke detection on CT scans, CAD systems automatically analyse the images and highlight regions of interest that may indicate the presence of stroke. By assisting radiologists in their interpretation, CAD systems reduce the likelihood of human error and aid in the identification of subtle abnormalities that may be missed during visual inspection. In fact, the role of CAD in ischemic stroke detection extends beyond assistance. CAD systems have the potential to improve patient outcomes by enabling early and accurate diagnosis, leading to timely intervention and treatment. By providing an additional set of "virtual eyes," CAD systems help radiologists detect stroke-related abnormalities that may be overlooked or misinterpreted, ultimately enhancing the overall accuracy of stroke diagnosis.

Various CAD systems developed to aid and assess physicians to provide accurate and fast detection and prediction of ischemic lesion. In fact, Current techniques such as CT, MRI protocols can help the experienced radiologist to detect ischemic changes especially in hyper- and acute stages, but if a general physician makes incorrect judgment, this causes missing of the best time for treatment. Therefore, it's important to enhance the quality of medical images for proper diagnosis and image recognition (Doi, 2007; Kasner, 2006).

In 2011, a purposed CAD scheme is published by Tang et al., using circular adaptive region of interest (CAROI) operated in B-NECT to assists emergency physicians and radiology residents and radiology specialist to detect early ischemic changes. B-NECTs were collected as DICOM format then preprocessed, by removing area that ischemic event impossible occur

as bone. After that, the brains aligned to the best symmetrical position. In the processing stage CAROI used to localize the regions in which the intensity changed on the improved CT images. The resulted output showing increased sensitivity and specificity of 93.33%, 90.3%. While there were variations in ROC carve depend on the observer's experience. Where the ROC of the emergency physicians and radiology resident improved (0.942 with CAD Vs. 0.879 without CAD), (0.990 with CAD Vs. 0.965 without CAD) retrospectively compared with radiology specialist ROC carve relatively remained constant (0.998 VS. 0.999) due to their experience (Tang et al., 2011).

Pedro et al., 2017 implemented a new approach to classify brain strokes of NECT through an algorithm used to analyze the brain tissue densities called Analysis of Brain Tissue Density (ABTD). The results are compared with multiple feature extractors such as Gray-Level Co-Occurrence Matrix (GLCM), Local Binary Pattern (LBP), Central Moment (CM), Statistical Moment (SM), Hu's Moment (HM), and Zernike's Moments (ZM). The superiority of ABDT is the shortest time extractor with higher average accuracy (99.3%) when it is combined with Optimum Path Forest (OPF) (Pedro P. Rebouc¸as Filho, 2017).

While Lo et al., in 2019, aimed to develop a CAD model to detect hyper-acute ischemic stroke lesions on 26 B-NECTs. They suggested depending on previous literature that the anatomical structures in the human brain are symmetric, thus means that when a stroke occurs, the ischemic lesion can be highlighted by making a comparison between the right and left-brain hemispheres. By tissue clustering techniques bone was removed. Then, the brain tissue is divided into right and left hemispheres for symmetric interpretation. After that, features were extracted by Ranklet-transformation and GLCM. For Ranklet- transformation, used to transfer the original pixels into relative contrast coefficient in the vertical, transverse, and diagonal orientation, and textural features extraction. Following this step, the images are divided into two blocks for comparison calculation. In relation to the GLCM matrix, for spatial variation in the gray level. The Co-Occurrence between adjacent pixels was calculated at (d) distance and multiple directions (θ= 0, 45, 90, and 134). 8 features were selected from Ranklet-transformation, 3 from GLCM then these features were combined. The results showed that there are better accuracy of Ranklet-transformation (81% Vs. 71%), sensitivity (64% Vs. 57%), specificity (90% Vs. 79%) in comparison with GLCM (Lo et al., 2019).

2.2.2 Machine Learning (ML)

ML, in other words, machine intelligence means the human brain mimicking by computer, is a major branch of AI, figure (7), and DL is a subtype of machine learning (Krittanawong et al., 2017).

This development of ML algorithms in medicine is related to the digital revolution that makes data collection and storing is more available and inexpensive (Mechaliski RS, 1998). This approach provides crucial tools for intelligent data analysis and diagnosis.

Broadly, ML mimics human thinking for tasks need human intelligence, by using statistical approaches enable computers to make optimal outcome decisions from large datasets used to train computers for pattern recognition of selected features extracted from these data to classify it into subtypes. Interestingly, ML can evaluate objects qualitatively and quantitatively. Moreover, ML characterized by the ability to detect subtle in voxel-level patterns, large scale implementation and speed (Goodfellow et al., 2016). Different ML algorithms were employed for the detection and prediction of stroke region, it improves the ability to detect the stroke lesion area within the brain.

On the other hand, Abedi et al., in 2017 aimed to develop a new tool to differentiate between acute ischemic stroke and stroke mimics by using a supervised learning method. Artificial neural network (ANN) models were used to learn and test stroke-like symptoms, and patient data, these data include patient history, hospital curse, discharge diagnosis, CTs, MRIs, and electroencephalography (ECG) records if available. SPSS was used to make statistically comparison between acute stroke and stroke mimics using square test (x^2). Results indicated that the sensitivity and specificity of ANN for the diagnosis of acute ischemic stroke was 80%

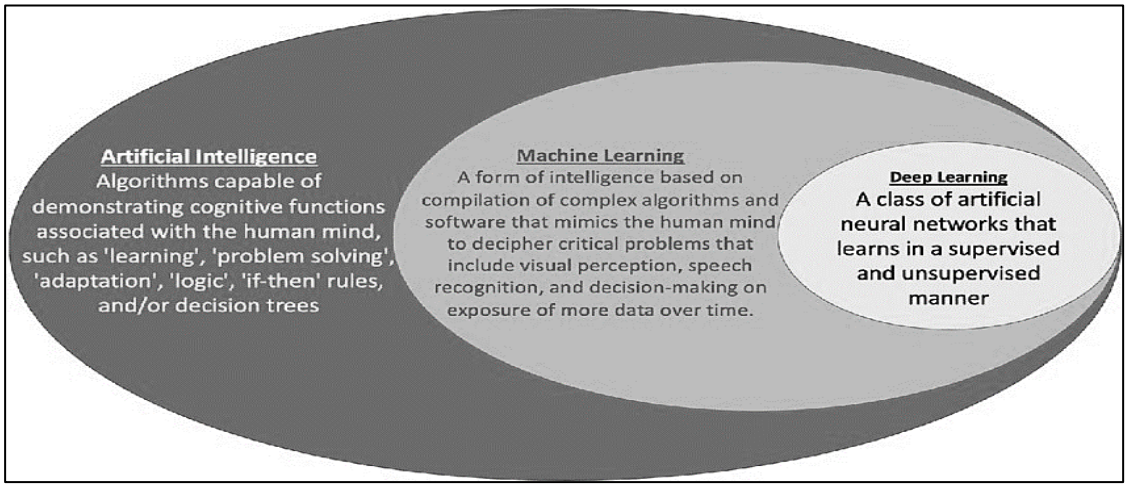


Figure 7: Artificial intelligence subtypes.

(95% confidence interval (CI) 71.8-86.3) and 86.2 (95% CI 73.4- 87.1) retrospectively and the median precision of ANN for acute ischemic stroke diagnosis was 92% (95%CI 88.7-95.3) (Abedi et al., 2017).

Roman et al., 2017 used the texture feature analysis to detect and localize the hyperacute ischemic stroke region. 139 B-NECTs for hyperacute ischemic stroke followed by DW-MRI. The authors firstly used the b-spline interpolation to match the spatial resolution of B-NECT and DW-MRI to register them geometrically by 3D symmetric diffeomorphic image registration to identify the brain hemispheres, and localizing the hyperacute stroke lesion. Then they identified the reference contralateral region which represent the normal area of the brain (ROIht) and its spatially corresponding stroke region (ROIs). After that texture analysis were done to extract and classify features of ischemic stroke and their contralateral region by Run-length matrix (RLM), also GLCM to characterize the spatial distribution of gray level. Finally, classification was done to the texture analyzed data (input) by non-linear SVM to determine the stroke lesion boundary, Decision Tree (DT) to determine tested data (decision nodes) stroke or health (leaf nodes). For more accuracy and to reduce the misclassification AdaBoost used. In this study, the longitudinal interhemispheric fissure is not exactly in the axial correspondence due to the head rotation and tilting. Poorly investigation for the data set of strokes from 0 to 2 hours from the symptom's onset. AUC for stroke 2-4.5 hours from the symptom's onset is 0.85 while between 4.5 hours to 8 hours 0.938-0.957 (Roman Peter. Panagiotis Korfiatis, 2017).

Interestingly, in 2018, Peixoto et al, reached the highest specificity of 99.1%, sensitivity of 97%, F1-score of 98% and accuracy of 98% for the classification task of 300 NECT of brain to normal, ischemic and haemorrhagic strokes (100 for each class) by Structural Co-Occurrence Matrix (SCM) when it compared with other feature extractors as LBP, Multi-Layer Perceptron (MLP), and the Least Square SVM (Peixoto & Rebouças Filho, 2018).

Kniep et al., 2020 proposed a ML-based approach to predict early ischemic changes depending on posterior circulation-ASPECT (pcASPECT). 552 B-NECT were registered and visually verified by two experienced clinicians, the region where stroke area demonstrated by pcASPECT were segmented added to standard maps, PyRadiomics Python package U2.1.0 is used to extract 1218 quantitative features from each pcASPECT areas. To improve the generalizability and reduce the bias resulted from unimportant features, 20 most important features were determined by Gini impurity. Then ML based algorithm (RF) performed to

predict early ischemic changes. Finally, the prediction performance of early ischemic changes was analyzed by AUC which were 0.7 for cerebellum and 0.82 for thalamus. Predictive performance of the classifier was significantly higher compared to visual reading for thalamus, midbrain, and pons (P value<0.05) (Kniep et al., 2020).

2.2.3 Deep Learning (DL)

Conventional machine learning techniques had their constraints when dealing with raw, unprocessed natural data. Creating a pattern recognition or machine learning system used to be a complex endeavor, requiring meticulous engineering and in-depth domain expertise. This involved the development of a feature extraction mechanism that converted the raw data, such as pixel values from an image, into a suitable internal representations or feature vectors. The learning subsystem, often a classifier, would then utilize this transformed data to recognize or classify patterns within the input (LeCun et al., 2015). representation learning operates as an artistic symphony, crafting multiple levels of representation by combining uncomplicated nonlinear modules. Each of these modules guides the transformation of raw input, initiating with the elemental and ascending to a more abstract stratum. Through the fusion of numerous such transformations, the feature learns intricate functions of remarkable complexity. In the task of classification, the upper layer of representation magnifies the essential discriminative aspects while diminishing the irrelevant fluctuations.

For instance, in DL, an array of pixels represents an image. The feature of this image is divided into layers. More clearly, for example, the first layer represents the presence or absence of sharp edges feature rearrangement. The second, represents the small motifs feature rearrangement, while the third layer rearranges the motifs in a large combination. In fact, the key aspect of DL lies in the fact that the learning is not designed by specialists, but it is designed from data itself using the proposed learning procedure (Clement Farabet et al., 2012; Krizhevsky et al., 2012; Mikolov et al., 2011).

DL provide advanced solution of major problems that faces AI for many years, due to the excellent ability to discover complex structures in high-dimensional data, so this approach is applicable in many domains as sciences, business and governmental procedures. Moreover, in speech and pattern recognition, especially in drug molecules activity, brain circuits analysis, DNA mutation prediction and detection and other diseases prediction and detection (Leung et al., 2014; Xiong et al., 2015).

2.2.3.1 From Supervised to Unsupervised Learning

Commonly, machine learning form is supervised learning. Suppose that we want to establish a system to classify images of dogs, cats, houses, persons. Firstly, we need a large data set of these mentioned type of images, then each category will be labelled as its class. In the training set, the machine is shown an image -for each category-, then vector of scores will be produced as output. Then, we need to compute an objective function to measure the error between the output scores and the desired pattern scores. After that -by machine- internal adjustable parameters as a real number called weights to reduce errors. These weights considered the check points that will define the input-output machine function. In a typical DL system, hundreds of millions of these adjustable weights.

Gradient vector, is a parameter to ensure that each weight in the proper value, by measuring the amount of error when the weight vector is increased or decreased in a tiny value. The objective function, when averaged across all training examples, resembles a rugged mountains within the vast expanse of high-dimensional weight values. The negative gradient vector serves as a guiding compass, pointing toward the steepest descent in this landscape, gradually leading it closer to a minimum point where the average output error is minimized.

In practice, the majority of practitioners utilize a technique known as stochastic gradient descent (SGD). This approach involves presenting the input vector for a few examples, calculating the outputs and errors, computing the average gradient for those examples, and adjusting the weights accordingly. This process is repeated for numerous small sets of examples from the training set until the average of the objective function no longer decreases. It is referred to as stochastic because each small set of examples provides a noisy estimate of the average gradient across all examples. Surprisingly, this simple procedure often discovers a good set of weights quickly, in comparison to more complex optimization techniques. Following the training phase, the system's performance is evaluated on a separate set of examples known as a test set (Bottou & Bousquet, 2007). The purpose of this is to evaluate the machine's ability to generalize - its capacity to generate meaningful responses to new inputs that were not encountered during training. In numerous real-world applications of machine learning, linear classifiers are commonly employed in conjunction with manually crafted features. A two-class linear classifier calculates a weighted sum of the components of the feature vector. If the weighted sum exceeds a certain threshold, the input is classified as belonging to a specific category.

The limitations of linear classifiers in effectively addressing complex tasks such as image and speech recognition have been recognized since the 1960s (Duda & Hart, 1973). Linear classifiers can only partition the input space into simple regions, specifically half-spaces separated by a hyperplane. However, these tasks necessitate an input-output function that remains robust to irrelevant variations in the input, such as changes in position, orientation, illumination, or speech pitch and accent. Simultaneously, the function should exhibit high sensitivity to subtle differences that are crucial for discrimination, such as distinguishing between a white wolf and a Samoyed, a wolf-like white dog breed.

At the pixel level, images of two Samoyeds in different poses and environments can exhibit significant dissimilarities, while images of a Samoyed and a wolf in the same position and similar backgrounds may appear quite similar. Linear classifiers, or other shallow classifiers operating directly on raw pixel data, are incapable of distinguishing the latter two scenarios while categorizing the former two together. This highlights the necessity for effective feature extraction methods that can address the selectivity-invariance dilemma. These methods should produce representations that selectively capture discriminative aspects of the image while remaining invariant to irrelevant factors like the animal's pose.

To enhance the discriminative power of classifiers, generic non-linear features, such as those employed in kernel methods, can be utilized (Schölkopf et al., 2002). However, these generic features, like those derived from the Gaussian kernel, often fail to generalize well beyond the training examples. Alternatively, the conventional approach involves manually designing effective feature extractors, which demands substantial engineering expertise and domain knowledge. Nevertheless, this challenge can be circumvented by leveraging the capabilities of deep learning, which enables the automatic learning of high-quality features through a general-purpose learning procedure (Bengio et al., 2005). This represents a significant advantage of deep learning over traditional methods.

In conclusion, the limitations of linear classifiers in handling complex tasks have motivated the exploration of more sophisticated approaches, such as deep learning, which can automatically learn powerful and discriminative features. By addressing the selectivity-invariance dilemma and enabling the extraction of meaningful representations, deep learning offers a promising avenue for advancing the performance of classifiers in various domains, including image and speech recognition.

researchers have long aimed to replace hand-engineered features with trainable multilayer networks in the field of pattern recognition. However, it was not until the mid-1980s that the solution of using multilayer architectures became widely understood. It was discovered that these architectures can be effectively trained using simple stochastic gradient descent, provided that the modules within the network are relatively smooth functions of their inputs and internal weights. The backpropagation procedure, which computes gradients using the chain rule for derivatives, emerged as a practical application for training multilayer networks (Rumelhart et al., 1986).

The key insight behind backpropagation is that the derivative (or gradient) of the objective function with respect to the input of a module can be computed by working backwards from the gradient with respect to the output of that module (or the input of the subsequent module). This realization was independently discovered by multiple research groups during the 1970s and 1980s.the development of trainable multilayer networks and the application of backpropagation have revolutionized the field of pattern recognition. These advancements have allowed for the automatic learning of features, replacing the need for manual feature engineering. The discovery of the backpropagation procedure and its practical implementation have paved the way for the widespread adoption of multilayer networks and have significantly contributed to the progress of machine learning and artificial intelligence.

Briefly, the backpropagation algorithm uses the chain rule of calculus to efficiently compute the gradients of the loss function with respect to the weights in a neural network. It works in two phases - forward and backward passes. In the forward pass, an input is fed through the network layer by layer to produce an output. The loss between this output and the true target value is calculated. In the backward pass, this loss value is propagated backwards from the final layer through the network. At each layer, the gradient of the loss with respect to the weights connecting the current and previous layers is calculated. This involves multiplying the gradient from the next layer by the derivative of the activation function of the current layer. Once the gradients for all weights are determined, an optimization algorithm such as stochastic gradient descent can use these gradients to update the weights, reducing the loss. By repeating this process over many iterations with different training examples, the weights are adjusted to minimize the loss across the entire training set.

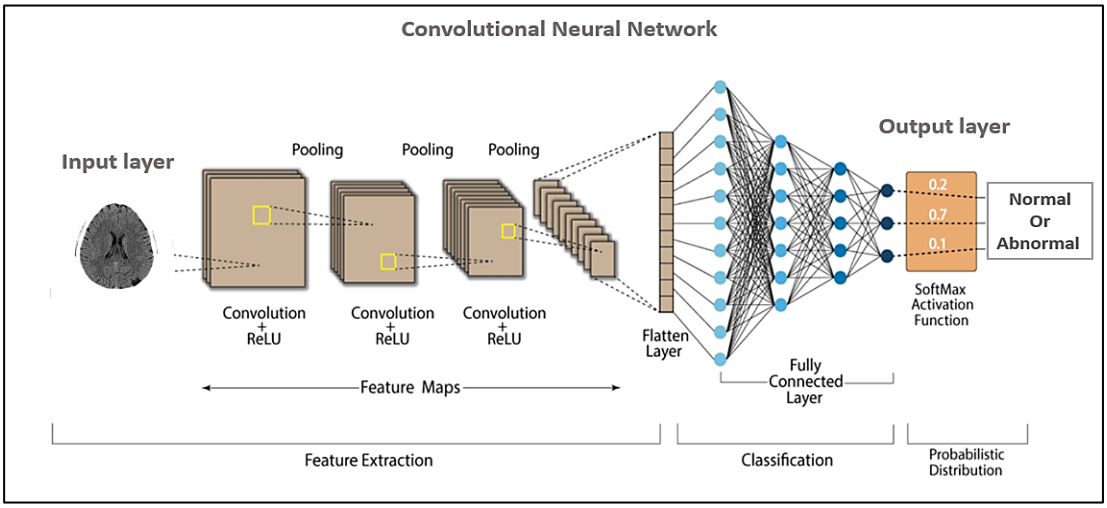
2.2.3.2 Convolutional Neural Networks (CNNs)

The convolutional neural network (CNN) proved to be a notable exception among deep feedforward networks in terms of trainability and generalization ability during a time when neural networks were largely disfavoured. CNNs achieve their success through the use of convolutional layers that utilize shared weights and local connectivity patterns between neurons in adjacent layers, inspired by biological processes in the visual cortex (LeCun et al., 1990; LeCun et al., 1998). This architecture allowed CNNs to be much easier to train than fully connected networks while also generalizing significantly better.

Figure 8: Illustration of convolutional neural networks (CNN).

Despite being developed when neural networks were out of favour, CNNs achieved many practical successes and were widely adopted by the computer vision community. Their recent resurgence can be attributed to their ability to learn hierarchical representations of visual data, leading to state-of-the-art performance on image classification and analysis tasks. CNNs can operate on input data represented as multiple arrays, including 1D arrays for signals/sequences like language, 2D arrays for images/spectrograms, and 3D arrays for video/volumes. Four fundamental concepts behind CNNs shown in Figure (8) utilize the nature of data. First, local connections between neurons allow CNNs to automatically learn filters to detect low-level features like edges or motifs. Second, shared weights enable these filters to identify patterns anywhere in the input. Third, pooling layers make the representation invariant to small variations by merging semantically similar features. Fourth, stacking multiple convolutional and pooling layers in a deep architecture enables CNNs to learn increasingly complex, hierarchical patterns composed of simpler elements. This replicates how higher-level concepts are formed from lower-level ones in visual processing and language.

Typical CNN models alternate between convolutional and pooling layers, followed by fully-connected layers to perform classification or regression (Cadieu et al., 2014). Backpropagation enables end-to-end training of the entire model. Early applications of CNNs achieved success in domains like speech recognition, document reading, and character recognition in the 1990s. Since then, CNNs have become the dominant approach for computer



vision and natural language processing tasks due to their ability to automatically learn discriminative feature representations directly from data (Waibel et al., 1989).

CNNs have achieved significant success in tackling computer vision problems that require processing of visual input data since the early 2000s. These models were shown to excel at tasks involving object detection, image segmentation, and recognition where sizable labelled datasets existed, such as for traffic signs, biological images, detecting faces, text, and human bodies (Clément Farabet et al., 2012; Hadsell et al., 2009; Sermanet et al., 2013; Simonyan & Zisserman, 2014). A notable application has been pixel-level semantic segmentation which holds promise for autonomous systems like robots and self-driving cars being developed by companies such as Mobileye and NVIDIA (Simonyan & Zisserman, 2014). While CNNs had gained traction for some problems, they were largely overlooked by the broader research community until a major breakthrough in 2012, when a CNN significantly outperformed competing methods on the large-scale ImageNet dataset, nearly halving error rates. This seminal work revived interest in deep learning approaches for computer vision and demonstrated that leveraging graphics processing units (GPUs), activation functions like rectified linear unit (ReLU), dropout regularization, and data augmentation techniques enabled training of CNNs with unprecedented scale, containing over 10 layers and hundreds of millions of parameters. Such deep CNN architectures are now the preeminent solution for the majority of visual recognition tasks, approaching and in some cases surpassing human-level performance (Krizhevsky et al., 2012).

Researchers first started using CNNs to look at medical images in 1993. They trained a CNN to find lung nodules (abnormal growths) in chest X-rays. The same group, Chan et al., then used CNNs to find microcalcifications in mammograms in 1993 (Lo et al., 1993).

Microcalcifications are tiny calcium deposits that could be a sign of breast cancer. They trained their CNN model to detect these. In the following year (1994-1995), Chan et al. applied CNNs to find masses in mammograms. Masses are abnormal tissue growths that also need to be identified, as they could be cancerous. They published their work on using CNNs for mass detection in mammograms in multiple papers between 1994-1995 (Lo et al., 1995; Sahiner et al., 1996; Wei et al., 1995).

Early studies showed CNNs could achieve expert-level performance in segmenting stroke lesions on MRI scans. Pereira et al. (2016) developed a deep CNN that segmented acute ischemic stroke lesions with a Dice coefficient of 0.88 compared to manual annotations by experts, demonstrating CNNs' potential to capture subtle lesions invisible to the naked eye This establishes CNNs as a promising approach for automated stroke diagnosis (Pereira et al., 2016).

Pan, in 2017, developed an automated system to detect ischemic changes on 256 pre-processed CT images called patch images as input to learn and test CNN-deep learning algorithm. The method of this paper is divided into four stages, pre-processing to remove any impossible stroke area as skull, or cerebrospinal fluid (CSF). Selection of a fixed-size of patch image as a second step to avoid the error in training steps that happened because stroke occur in different area. The third step is data augmentation for increasing the number of data to avoid overfitting problems. Finally, the patch images were used as input to train and test CNN. The authors concluded that the accuracy reached from this system is 90% which can effectively assist the doctor's diagnosis (Pan, 2017).

Lisowska et al., in 2017 investigated the use of contextual information for the detection of stroke signs by using 3D CNN for unsupervised bilateral comparison of the symmetry between paired points in both right and left hemispheres. this model is inspired by the workflow of radiologists. Selected image patches were extracted from both hemispheres to parallel CNN channels to allow right and left comparison from the aligned volume of B-NECTs. Moreover, the anatomical context of brain tissue was compared with a normative atlas created from normal examples by 3D CNN. For training, of the dense vessel and ischemia at the voxel level, of 71 datasets. The prediction scores above zero (positive) indicate a stroke. If the prediction score is closer to the zero then stroke prediction is smaller, while the negative score indicates normal tissue. Finally, the model performance was evaluated by Titan GPU. Results showed that the ROC, and AUC of the model in ischemic detection were (0.915, 0.783) retrospectively in the case of CNN with atlas, which is nearly same with bilateral CNN comparison only. For

dense vessel detection, bilateral CNN with atlas showed better ROC, AUC (0964, 0.898 retrospectively) compared with the use of bilateral CNN only, ROC = 0.891, AUC = 0.691 (Lisowska et al., 2017).

More recent works continue leveraging CNNs for automated stroke diagnosis. Acharya et al. (2018) reviewed promising results of CNNs for stroke lesion detection and segmentation across modalities. Also, discussed future directions like multimodal data integration and explainable AI to increase clinician trust in CNN-based systems (Acharya et al., 2018).

Subsequent studies applied CNNs to other stroke-related tasks and modalities. Liu et al. (2019) trained a CNN on CT perfusion images to detect acute ischemic strokes with over 96% accuracy. Li et al. (2018) created a multi-scale CNN to jointly detect and segment hemorrhagic and ischemic strokes on head CT scans, achieving Dice coefficients of 0.81-0.85 (Li et al., 2018).

In 2019, Clèrigues et al. aimed to evaluate an automated DL method called Ischemic Stroke Lesion Segmentation (ISLES) which used to extract potentially quantifiable information. In other words, to detect ischemic core in CT and CTP images. 94 cases were included for training set, while 62 cases were used for testing set. A 2D patch-based DL approach was used in the training set. However, the detection and segmentation of core in all images is less than 5% of the brain tissue, so an issue called class imbalance which leads to misclassification due to the significantly fewer examples in the training set in one or more classes. This issue has been processed by data augmentation, dropout layers that deactivates or sets the noisy updates to zero, early stopping which is a technique that halts the training process when further acquisition of generalizable knowledge is no longer possible. Finally, by made a combination of the uncertain classification estimations then use the highly overlapping patches to reduce the segmentation artifacts. Cost-sensitive loss functions were used for the imbalanced classes which modify the standard loss function in ML and DL. 2D (Clèrigues et al., 2019)

In 2022, Syu-Jyun et al., reached to detection accuracy of 93.9% in the detection of acute stroke lesions using 59 B-NECT and 59 DW-MRI images which underwent pre-processing steps to improve the contrast of brain tissue, brain alignment, and orientation adjustment and normalization of B-NECT and DW-MRI images spatially to the Montreal Neurological Institute (MNI) brain CT and brain MRI templates. After pre-processing steps, t-score mapping for infracted and normal tissue was created for each CT and MRI. Then t-score

maps were used to train 16*16 matrix CNN that including 17 layers. the sensitivity and specificity of the system were 98.4%, and 89.8% respectively (Syu-Jyun Peng, 2022).

Recently, DL has created a significant impact in the medical field, especially for nonuniform parameter relationships several researches aimed to detect and segment stroke lesions automatically without supervision. Representation learning comprises a collection of techniques that enable a machine to ingest raw data and automatically discover the essential representations necessary for tasks such as detection or classification (Lebedev et al., 2020). While transfer learning, is a process where the features are extracted from a machine learning technique, then these features are reused as a starting point for a model on a target task. This method requires less labeled data in the target domain compared to training models from scratch, which often requires vast amounts of annotated data. This makes transfer learning applicable even when limited target data is available (Pan et al., 2010).

However, a lack of very large annotated medical imaging datasets remained a challenge for CNN training. Transfer learning techniques were explored to address this. Kooi et al. (2017) showed fine-tuning a CNN pre-trained on the natural image dataset ImageNet improved stroke lesion detection over training from scratch with limited medical data (Kooi et al., 2017).

Wu Qiu et al. published in 2020, a paper on detection of early infraction in acute stroke in B-NECT. The paper presents an automated ML approach for quantification and detection of acute ischemic stroke (AIS) and compared the output to DW-MRI as a reference image. 257 patients had AIS (<6 hours from the symptoms onset), 157 for training and 100 for testing. B-NECT and DW-MRI were done for each patient. Manual segmentation of the ischemic lesion for DWI (ITK-SNAP), volume computed, skull tripped and registered by Montreal Neurologic Institution 12. CNN used to extract 5 features which are the mean Hounsfield units, the density differences in the bilateral symmetric brain regions, the hypoattenuation measurements, the degree of hypoattenuation, the distance feature and the atlas-encoded lesion location feature. These 5 features are applied to train the random forest classifier. The residual 100 images used to test the classifier. 11ml is the resulting difference between the algorithm output (CT images) and DW-MRI, P = 0.89 (Wu Qiu, 2020)

Other publications designed hybrid models between ML and DL approaches as Jun-Lu et al. in July of 2022, developed a DL model compromised of two deep CNN one for localization of acute ischemic stroke and the other for AIS classification on B-NECT for 1136

suspected non-visible AIS. Depend on DW-MRI the non-visible AIS is labeled by two third-year radiology residents. Both labeled AIS B-NECT and negative AIS B-NECTs were used as input to train the first localizer model called YOLO v3 in which were the output is a cropped rectangular area with AIS probability. The output then used as input to the second deep CNN classifier model called 50-layer Residual Network (ResNet 50) its output represented the diagnostic probability of AIS. The performance of the model was evaluated by the area under the receiver operator characteristic curve (AUC), sensitivity, specificity and accuracy values. 83.6% is the AUC of the model with sensitivity 67%, specificity 98.2% and 89.9% accuracy (Jun Lu, 2022).

2.3 Litrature Review Summary

Neuroimaging is essential for accurately diagnosing acute ischemic strokes, distinguishing them from other brain conditions like hemorrhages or tumors, and showing affected brain areas. The most commonly used tool is the brain non-enhanced CT (B-NECT), known for its wide availability and cost-effectiveness. However, its ability to detect ischemic changes soon after a stroke occurs is limited, improving significantly only after the first day. Efforts to enhance CT's effectiveness, such as adjusting imaging settings, have shown that better detection can sometimes reduce clarity or specificity.

Recent improvements highlight Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) as more sensitive than B-NECT, although it is slower and can be affected by patient movement. CT Perfusion (CTP) scans, which provide a dynamic view of blood flow in the brain, have proven useful in diagnosing ischemic strokes by showing how blood moves through the brain’s capillaries.

The Alberta Stroke Program Early CT Score (ASPECTS) is another advancement, offering a way to quickly assess the severity of a stroke, which can guide the decision on whether to perform surgery. This tool highlights the importance of quick and precise imaging in treating strokes.

The use of AI, especially through Computer-Aided Diagnosis (CAD) systems, shows a significant advance in stroke diagnosis. These systems improve image quality, identify critical features, and use machine learning to enhance diagnosis accuracy. CAD helps detect subtle signs of strokes that might be missed by humans, reducing errors and improving treatment.

Moreover, Deep Learning (DL) and Machine Learning (ML) have further enhanced CAD's capabilities. Convolutional neural networks (CNNs), a type of DL, are particularly effective for analyzing complex image data. These techniques have shown high accuracy in distinguishing between different types of strokes on CT scans and have often outperformed traditional methods.

Recent studies have used advanced ML and DL models to improve the detection and classification of stroke signs, achieving higher sensitivity and specificity. Some research combines these models to enhance both the identification and categorization of stroke-related abnormalities.

Our study introduced a novel approach for improving the efficiency of ML in predicting Ischemic stroke diseases called SPEM model. The SPEM model leverages recent technological advancements by integrating Contrast Limited Adaptive Histogram Equalization (CLAHE) for image pre-processing. This method improves the visual quality of B-NECT images, making it easier to identify ischemic areas by enhancing contrast. This pre-processing step is crucial as it addresses the inherent limitations of B-NECT, particularly its lower sensitivity in detecting early ischemic changes shortly after stroke onset.

Furthermore, SPEM utilizes Densely Connected Convolutional Networks (DenseNet-121) for feature extraction. This deep learning approach is adept at processing complex image data, capturing subtle nuances that traditional methods might miss. DenseNet-121 is known for its efficiency in image classification tasks due to its ability to reuse features, making it highly suitable for medical imaging where precision is critical.

For classification, SPEM employs a combination of advanced machine learning classifiers: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). This hybrid approach allows the model to evaluate which classifier performs best under different conditions, optimizing accuracy, precision, and speed. Notably, our results indicate that the Logistic Regression classifier, when used in conjunction with DenseNet-121, provides the highest performance metrics across various stages of ischemic stroke, particularly in the hyper-acute stage.

The Stroke Precision Enhancement Model (SPEM) makes significant contributions to the existing literature by addressing key challenges in stroke neuroimaging. Firstly, it enhances the early detection of ischemic strokes, a crucial period where rapid diagnosis can dramatically

affect treatment outcomes. This is very important considering the limitations of current imaging methods in detecting subtle changes shortly after a stroke occurs. SPEM also supports real-time diagnostic applications by optimizing processing time, marking a notable advancement over traditional, slower methods that are less suited for emergency settings. Additionally, the integration of advanced preprocessing and classification techniques ensures that SPEM achieves high accuracy and reliability, outperforming traditional imaging methods and providing clinicians with dependable diagnostic information. Finally, the model's design allows for easy adaptation and scalability across various clinical settings, including those with limited resources, addressing common challenges in global health contexts.

Chapter Three: Methodology

This chapter describes the methodology adopted by this study to achieve the aim and objectives stated in sections 1.4 and 1.5 in Chapter 1. Section 3.1 discusses the dataset type and sources, dataset size, and scan techniques used to create data; section 3.2 explains the pre-processing techniques used to prepare the dataset for training and testing steps; 3.3 and 3.4 describe the DL and ML techniques used to build up and train the models. Finally, section 3.5 discusses the testing step used to evaluate the performance of the models.

3.1 Dataset

3.1.1 Data collection and data sources

The Research Ethics Committee of Al-Quds University approved this study. This analytical study involves a retrospective collection of medical imaging data acquired over 24 months from January 2021 through January 2023. Reported Dicom Imaging and Communication in Medicine (DICOM) files were extracted from the Paicture Archiving and Communication

(PACS) from the Radiology department of three hospitals located in the West Bank of Palestine: Specialized Arab Hospital (SAH) and An-Najah National University Hospital (NNUH), both located in Nablus city, and Ibn Sina Arab Hospital (IAH) located in Jenin city. The data comprised B-NECT scans for all cases and DW-MRI with apparent diffusion coefficient magnetic resonance imaging (ADC-MRI) scans from patients presenting with hyperacute, acute ischemic stroke, and normal cases within 3 months. Experienced radiology specialists report all retrospective imaging data gathered.

3.1.2 Data Selection

Our data targeted all ischemic stroke patients administered to these mentioned hospitals between 01/2021 and 01/2023. So, the inclusion criteria for this study were B-NECT for both the normal group and all stroke patient categories. DW/ADC-MRI for normal group, hyper-acute, and acute stages for stroke patients. All corresponding collected imaging data were reported in this study. Exclusion criteria: patients with brain malformations or pathologies such as tumors, cysts, abscesses, hemorrhages, or other diseases. Images with severe artifacts (metallic implants or surgical clips, motion, etc.). Unreported CT or MRI images. For acute and hyper-acute stages, patients with missing CT or MRI data were excluded.

3.1.3 Slices Selection

Depending on radiology specialist reports, the data were classified into normal and stroke groups. For stroke groups data were categorized as reported according to time from symptom onset into hyperacute (<6 hours), acute (6-24 hours), subacute (1-7 days), and chronic (>7 days) CVA. All collected images in the DICOM format were first converted to Joint Photographic Experts Group (JPEG) format; for analysis. Slices with ischemic stroke lesions were selected in all categories. For patients in hyper-acute and acute stages the slice selection and lesion identification were performed by using DW/ADC-MRI images as reference; to pinpoint the restricted regions in DW-MRI that correspond to hyper-acute and acute lesions. By this processing workflow, the study ensured a lesion-focused slice selection and standardized the image formats.

3.2.3 Data size

In this study, 2815 images included various patient outcomes: 1025 normal images, 315 hyper-acute stroke images, 1000 acute stroke images, 225 sub-acute images, and 250 chronic stroke images.

3.2.4 Scanners Parameters and Desktop Characteristics

Three CT and three MRI scanners were included in this study. These scanners are installed in the radiology departments of the three mentioned private hospitals in WB of Palestine. Table (1) summarized the scanners types and the scanning parameters used to acquire images.

Table 1: Scanner types and scanning parameters included in this study.

Hospital name	**IAH**	**SAH**	**ANNUH**
Brain CT			
Product name	Philips incisive	Philips incisive	Toshiba (Canon)
Scanner type	128 slices	128 slices	168 slices
Tube current (mA)	300	280	300
Peak voltage (kVp)	120	120	120
WW/ WL	80/ 40 HU	80/ 40 HU	80/ 40 HU
Matrix size (mm)	512*512	512*512	512*512
Slice thickness (mm)	0.625	1	0.5
Reconstruction technique	iDose 3	iDose 3	AIDR 3D Integrated
Brain DW-MRI sequence			
Product name	Philips Inginia 1.5 T	Philips Inginia 1.5 T	Philips Inginia 1.5 T
TE (ms)	85	90	87
TR (ms)	4000	3354	3710
FOV (mm)	230*230	230*230	230*250
Voxel size	1.5*2.2*5	1.6*2.53*5	1.5*2.2*5
Matrix	152*106	144*90	152*106
b-Factor	0&1000	0&1000	0&800

The study was conducted using Python 3.9.18 as the primary programming language. Anaconda environment was set up to manage packages and dependencies. Preprocessing routines, model architectures, training loops, and performance metrics were all coded and run natively in Python on a desktop computer with the following specifications: Intel 12th Gen Core i7-12700K CPU, 32GB RAM, and NVIDIA GeForce RTX 4070 Ti 12GB GPU. This allowed for computationally intensive tasks like model training and hyperparameter tuning to be performed efficiently. The unified Python workflow, from data handling to results analysis, was made possible using these specifications and ensured reproducibility.

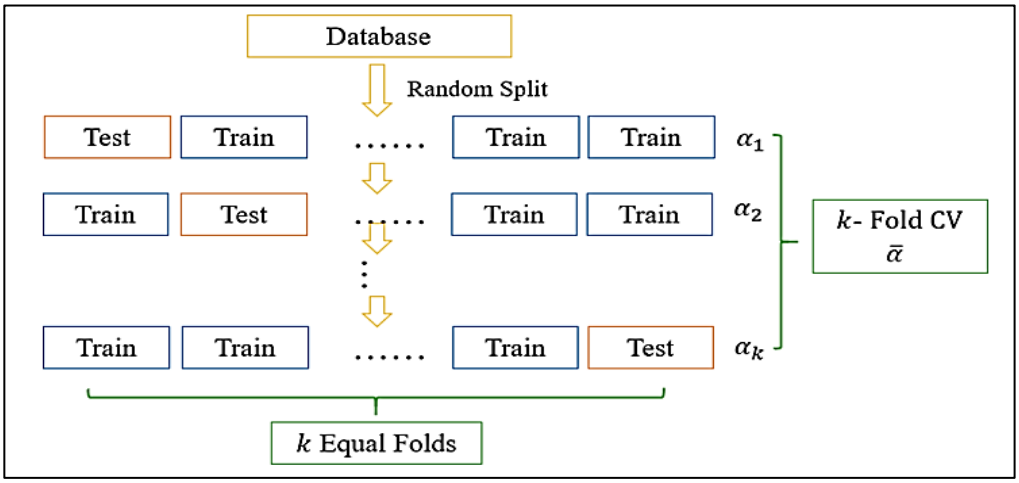


Figure 9: The main architecture of k-fold cross-validation. Copyright 2020 Elsevier

3.2.5 Cross-validation

Cross-validation is an important technique for evaluating and comparing machine learning models to reduce problems like overfitting on a particular dataset split. In this study, a 25-fold cross-validation approach was adopted to rigorously assess classifier performance. The dataset was divided into 25 equal partitions, with 24 folds used for training and the remaining fold for testing. This process was repeated 25 times so that each partition was used once for testing. Taking the average of the results from the 25 iterations helped produce a more robust and reliable estimate of classifier generalization capability compared to a simple train-test split. Using such a large number of folds (25) also helped maximize the data available for both training and validation purposes. The cross-validation results provided an unbiased evaluation of how well each model and preprocessing technique would perform on unseen data.

3.2 Dataset Preparation and Pre-processing.

3.2.1 Data Sorting

The prediction model was implemented into four data set groups of binary classification problems: (hyper-acute/ normal, acute/ normal, sub-acute/ normal, and chronic/ normal). 80% of the dataset was used for training, and 20% for testing the model performance.

3.2.2 Data Manipulation

In this study, all B-NECT images were resized from their original resolution of 512x512 pixels to 640x640 pixels through bilinear interpolation. This preprocessing step was necessary to

standardize the input dimensions required by the feature extractor (DenseNet-121). Resizing to a larger, fixed resolution allowed all images to be processed consistently by the deep learning architecture during training, despite varying native resolutions in the CT images.

One of the important data manipulation steps is the removal of irrelevant information that lead to increase the computational time and degrading the model performance, as image cropping and undesirable anatomic structures removal (Salvi et al., 2021). In this study, in addition to image cropping, by Python-based ANTs toolkit a semiautomated technique was used to extract brain tissues while skull excluded via atlas-based registration to mask brain. This process will focus the model in the brain parenchyma and reduce the irrelevant information, and thus minimize the probability of misclassification and computational process.

Moreover, all B-NECT underwent filtering processing including Gaussian filtering and Unsharp masking Gaussian filtering was applied to reduce noise and high-frequency artifacts present in the raw images. This smoothing step helped remove noise and fine details that could potentially confuse the model during training.

However, some loss of edges and textures was also introduced. To counteract this as reported by Kansal et al. in 2018, Unsharp masking was then performed on the Gaussian-filtered images that served to enhance and emphasize edges and boundaries between tissues or regions of interest (Kansal et al., 2018). This processing steps aimed to improve image quality and interpretability for the model by reducing noise while retaining important edges and textures that aid in tissue differentiation and ischemic lesion detection on CT images.

3.2.3 Stroke Precision Enhancement Model

Raw medical images often contain noise, artifacts and other anomalies that degrade the diagnostic and analytical quality (Zhang et al., 2001). Image enhancement play a pivotal role in the improvement of visual appearance and the analytical quality. With the advances of AI especially deep learning, image enhancement has gained increased importance (Rani & Kumar, 2013).

The priority of image enhancement before AI modeling is to optimize analysis and maximize performance. In fact, deep learning requires large, high-quality datasets to identify complex patterns, but noise and inconsistencies in raw images introduce biases that mislead models during training. Enhancement reduces elements and standardizes of visual

representation for automated pattern recognition (Anthimopoulos et al., 2016). For this study, the aim is to build an accurate stroke prediction and classification model, so enhancing the relatively low contrast of CT images due to the large dynamic range of CT numbers that affect the delineation of the interested region for abnormal area that represent ischemic stroke, and increasing the visual appearance and analysis quality is very important by applying an appropriate contrast enhancement technique (Tan et al., 2012).

Image histogram represent the distribution of gray values in digital images. For an image when the histogram skewed to the right side indicates a bright image, and if the histogram is skewed in the left side, it indicates a dark image, so the ideal image should be well distributed across the different grayscale levels. Histogram equalization (HE) algorithms is higher contrast manipulation by transformation spreads out the image intensity values along the total range [0,1] (Krutsch & Tenorio, 2011).

Contrast-limited adaptive histogram equalization (CLAHE) is one of the pixels-based division techniques, a spatial domain process that is used to improve small gray-level regions. The main idea of this technique as described by Reza AM in 2004 is to calculate histograms locally for different regions of the image rather than globally. This allows for adaptation based on neighboring pixel values by first, dividing the image into non-overlapped equal regions then the result is three groups: corner regions (CR), border regions (BR), and inner regions (IR). Secondly, each region will be represented by a histogram. Thirdly, the contrast is limited by a specified clip limit, then the clipped histograms are used to calculate cumulative distribution functions (CDFs) which show the proportion of pixels below each intensity value.

To determine the mapped intensity value for each pixel in the output image, CLAHE identifies the pixel's four closest neighboring regions based on a grid partitioning of the input image. The original intensity value of the pixel is then used as an index to lookup the grayscale mappings for that intensity at each of the four neighboring regions, as defined by their respective CDFs generated during histogram equalization. Bilinear interpolation is applied to these four mappings to derive the transformed value for the current pixel location. Finally, this mapped intensity is rescaled to fall within the output range from minimum to maximum and assigned to the corresponding position in the enhanced output image (Reza, 2004).

In this thesis, the CLAHE technique was employed; due to the ability of CLAHE to improve the low-contrast medical image and artifacts minimization, thus, standardizing data

representation and focusing the model on the aimed clinically relevant features, the output image named Stroke Precision Enhancement Model (SPEM).

To determine the optimal clip limit parameter for CLAHE, systematic evaluations were conducted using different values on a random sample of 500 images from the dataset. by applying CLAHE with a low clip limit of 2, then the clip limit was increased gradually at 6 levels (2, 2.5, 3, 4, 8, and 20) and re-assessed the sample images. For basic evaluation of the effectiveness of CLAHE within SPEM, quantitative assessment methods, such as the average Effective Measure of Enhancement (EME) and average Peak Signal to Noise Ratio (PSNR), are used. The EME and PSNR help to estimate the degree of improvement in image quality and the fidelity of the enhanced image to the original CT scan. These measures ensure that SPEM not only enhances the image for visual interpretation but also maintains the integrity and accuracy necessary for reliable stroke prediction.

The following equation is utilized for EME:

$$EME = \frac{1}{K_1 K_2} \sum_{L=1}^{K_2} \sum_{K=1}^{K_1} 20\, log\, \left(\frac{I_{max}(k,l)}{I_{min}(k,l)}\right) \quad (1)$$

Where K_1, K_2 are the number of horizontal and vertical blocks in the image, $I_{max}(k,l)$, and $I_{min}(k,l)$ are the maximum and minimum pixel values in each block.

PSNR was used to measure the deviation of the current image from the original image with respect to the peak value of the gray level. Given a reference image f and a test image g, both of size $M \times N$, the PSNR between f and g is defined by:

$$PSNR(f,g) = 10(255^2/MSE(f,g)) \quad (2)$$

Where MSE, is mean squared error

$$MSE(f,g) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(f_{ij} - g_{ij}\right)^2 \quad (3)$$

Within SPEM, as the MSE diminishes towards zero, the PSNR escalates towards infinity. This inverse relationship is pivotal, indicating that a surge in PSNR is synonymous with enhanced image quality. Essentially, a higher PSNR in the SPEM model signifies an image that more

closely represents the original CT scan in terms of quality, ensuring that the enhanced image retains the necessary details for accurate stroke prediction.

3.2.4 Data Augmentation

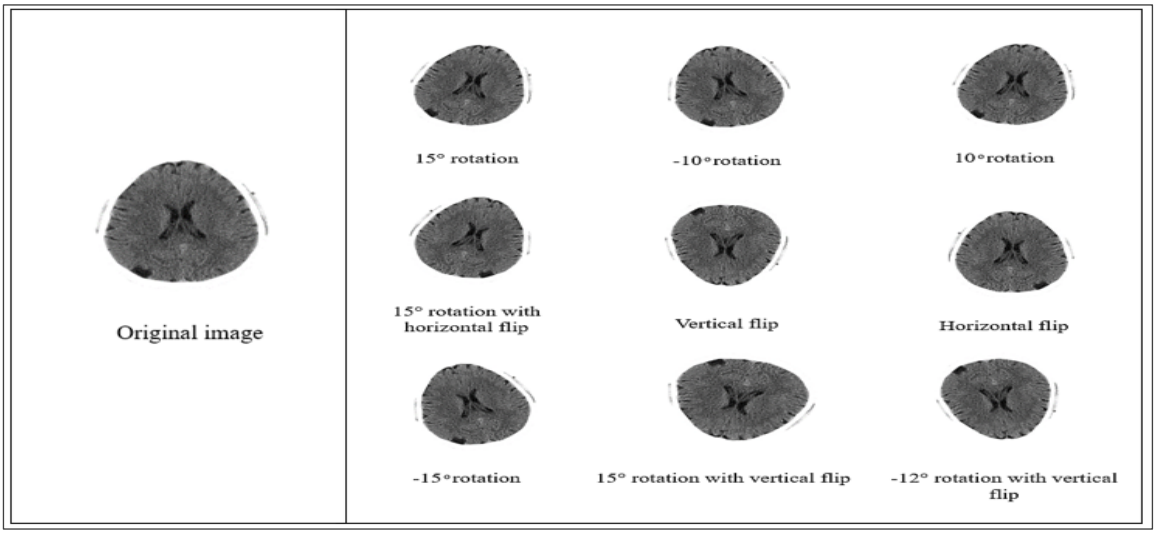


Figure 10: Data augmentation techniques employed in this study, the augmentation process involves rotation (-15° to +15°), horizontal and vertical flipping

One commonly employed strategy pre-processing technique in deep learning is data augmentation, which expands the limited medical imaging dataset and avoids overfitting. Overfitting occurs when the model is learned specific details and limited features as a result the lack of dataset used in training process, this leads to make the model is too focused on the specifics of the original training examples instead of the general patterns. And so, inaccurate output of new example (Krizhevsky et al., 2012).

In the current study, the training data was augmented using various geometric and photometric transformations to generate more robust training data for the models as shown in the Figure (11). Specifically, images underwent random vertical and horizontal flipping, and rotations between -15 to 15 degrees. The augmented images were added to the original dataset, hence the dataset increased by factor of 4 (11,260 images).

3.3 Deep Learning Feature Extraction (DenseNet-121).

Densely Connected Convolutional Networks (DenseNet) is a deep convolutional neural network architecture developed by Huang et al. in 2017 for computer vision tasks including image classification, segmentation and object detection. It was a significant improvement over

previous CNNs due to its dense connectivity which improved feature propagation and mitigated the vanishing gradient problem (Huang et al., 2017).

Vanishing gradient "wash out" problem common issue encountered when training deep neural networks using gradient-based optimization methods that used to minimizing the loss function (i.e., is function to measure the error rate between the prediction from layer to the actual target). In general, vanishing problem occur in deep networks with many stacked layers, the gradients tend to get smaller and approach zero as they are backpropagated from the final layer to the earlier layers. As the gradients get smaller moving backwards through many layers, they eventually become so tiny that they are rounded to zero and do not provide enough signal for the early layers to update their weights effectively through gradient descent. This causes the early layers to learn very slowly or get stuck. It makes training deep networks very difficult.

In CNNs, each layer is only connected to the subsequent layer. However, in DenseNets each layer is connected to every other layer in a feed-forward fashion. This "dense connectivity" facilitates feature reuse and strengthens feature propagation from earlier to later layers. It led to improved parameter efficiency over traditional CNNs.

DenseNet-121 has 121 layers and contains dense blocks where each layer is connected to all

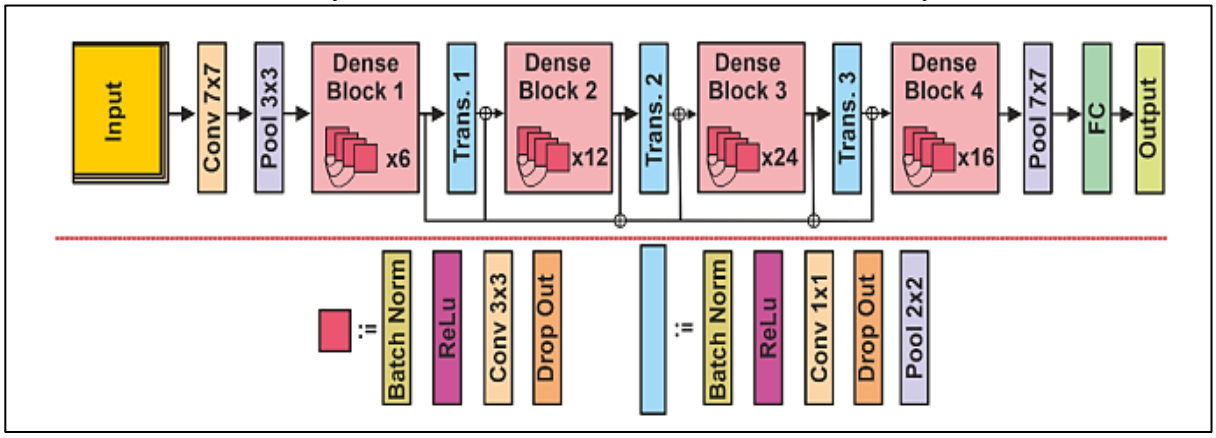


Figure 11: A schematic illustration of the DenseNet-121 architecture (Unluturk 2023).

subsequent layers. Within each dense block, the feature-maps of all preceding layers are used as inputs, enabling additional feature reuse. The first dense block contains 6 layers, while subsequent blocks contain increasingly more layers (12, 24, and 16 layers respectively) Transition layers between dense blocks implement bottleneck layers using 1x1 convolutions for dimensionality reduction and feed into subsequent dense blocks (Huang et al., 2017), figure(11).

DenseNet121 takes a raw input image and passes it through the initial input layer. Furthermore, the input layer performs an initial 3x3 convolution on the image, and subsequently performs 2x2 max pooling. Consequently, this creates the first set of feature maps from the input image that are then passed to subsequent layers. Moreover, the initial layer acts to extract basic features from the raw image data.

The key component of DenseNet-121 is that each layer receives the feature-maps of all preceding layers as input through concatenation. If the l-th layer generates m feature-maps, and k denotes the growth rate of the network, then the l+ l-th layer receives a total of m+k feature-maps as input. This concatenation can be represented by the equation:

$$x\,l\ +1\ =\ [x\,l, H\,l] \quad (4)$$

Where x l+1 is the input to the l +1-th layer, x l is the output of the previous l -th layer, and H l is the feature-maps generated by the l -th layer itself. This concatenation allows each layer to access the collective knowledge of all preceding layers. It encourages feature reuse and leads to more accurate and more efficient models. The network, therefore, exhibits a very fast convergence during training.

$$x`\ H`([x0, x1, \ldots, x`1]) \quad (5)$$

DenseNet-121 has a growth rate (k) of 32 and a compression factor of 0.5. This means each layer produces k=32 additional feature-maps.

Additionally, DenseNet121 contains four dense blocks which, in turn, make up the bulk of the network architecture. Each dense block is composed of bottleneck layers that perform composite functions of batch normalization, ReLU activation, 1x1 convolution to reduce the feature space, followed by more batch normalization, ReLU and 3x3 convolution. Consequently, this bottleneck structure helps reduce computation costs while simultaneously promoting feature reuse and propagation between layers. Furthermore, a key component of dense blocks is the concatenation of feature maps. Specifically, each layer in a dense block receives the feature maps of all preceding layers as input through concatenation. This encourages collective sharing of knowledge learned at each step.

Furthermore, between each dense block are transition layers. Transition layers perform the functions of bottleneck layers through 1x1 convolution, as well as 2x2 average pooling. This

serves to reduce the feature space and prevent overfitting after each dense block. Additionally, transition layers also help transfer the learned features to the subsequent dense block. Finally, after passing through the four dense blocks and three transition layers, the final set of feature maps undergo global average pooling. This helps aggregate the features. The pooled features are then passed to a fully-connected classification layer, where the images are ultimately classified into different categories based on the hierarchical representations learned from the input to the final layer (Huang et al., 2017).

DenseNets-121, as convolutional neural networks, are capable of extracting various levels of visual features from images due to their dense connectivity. At the lowest level, early DenseNet-121 layers can capture basic texture patterns, colors and edges through their small receptive fields. This is due to the application of convolutional operations with limited scope. Subsequently, mid-level features are derived as subsequent layers combine these low-level elements to detect more complex visual patterns characteristic of object parts. By virtue of increased network depth, DenseNets-121 can identify mid-level motifs that compose object sections (Huang et al., 2017; Li et al., 2017).

Deeper layers within DenseNets-121 then develop high-level semantic representations for full objects and scenes. Through composite function learning across layers, DenseNets gain the ability to recognize entire objects and environments useful for classification/detection tasks (Huang et al., 2017; LeCun et al., 2015). Moreover, the convolutional structure of DenseNets allows them to encode spatial relationships between extracted low/mid-level elements. This enables DenseNets to represent the arrangement of visual components within images (Huang et al., 2017; Wang et al., 2013)

Subsequent studies applied DenseNet121 to various medical imaging tasks. Chen in 2016 used DenseNet121 for brain tumor segmentation from MRI scans, outperforming prior methods (Chen et al., 2016). While, Jegou in 2017 used it for skin lesion classification from dermoscopic images, achieving 96.4% accuracy (Jégou et al., 2017). Also, DenseNet-121 was applied for lung nodule classification on CT scans, excellent accuracy was achieved of 95.6% (Li, 2018).

In this study, feature extraction was performed using DenseNet-121 that implemented in Python in both SPEM and CTs without preprocessing separately. As a deep-CNN pre-trained network structure with 121 layers arranged in a dense block connectivity pattern, DenseNet-

121 utilized to extract high-level image features from datasets. Both datasets were resized to 640 x 640 pixels, then passed through the DenseNet-121 model with a growth rate of 32. Specifically, in this study, 1024 features were extracted from the final pooling layer before the classifier head of the default DenseNet-121 architecture for each image. This pooling layer aggregates the feature maps into a 1024-dimensional vector for each image. In this way, 1024-dimensional feature vectors representing anatomical and pathological patterns were extracted from each image in the brain CT dataset using the default DenseNet-121 parameters. These extracted features were then used for training classifiers in this study to facilitate transfer learning from ImageNet for stroke classification.

3.4 Machine Learning Classification

classification is a core machine learning task involving the prediction of categorial labels from data examples. The goal of classification algorithms is to learn discriminative patterns from labeled training instances that enable the accurate classification of new observations. Supervised classification techniques utilize examples annotated with target class attributes to build predictive models capturing the underlying relationships between features and classes (Alpaydin, 2020). Common approaches for classification include decision trees which partition the feature space recursively (Quinlan, 1986), logistic regression which models class probabilities using logistic functions (Hosmer Jr et al., 2013), and support vector machines which find an optimal separating hyperplane in potentially high-dimensional space (Cortes & Vapnik, 1995). The performance of these algorithms is typically evaluated using metrics such as accuracy, precision, recall, and F1 score to quantify successful classification (Powers, 2020). Classification plays a vital role across diverse domains by enabling automated systems to predict categorical labels for new data based on learned classification models (Kotsiantis et al., 2007).

The fundamental task aimed in this study, is to evaluate which of this three widely used machine learning classifiers: Random Forrest (RF), Logistic Regression (LR) or Support Vector Machines (SVM) is better performance to categories the extracted features from DenseNet-121 into 5 classes (normal, hyper-acute, acute, sub-acute and chronic).

3.4.1 Random Forest (RF)

Decision tree (DT) which is a simple nonparametric supervised machine learning method purposed by Breiman in 2001, for solving classification and regression problems by partitioning the features and fitting a simple prediction for each part in the feature space.

The tree is constructed in a top-down manner. It builds a tree structure where internal nodes represent features in the data and branches represent different values or thresholds for those features as shown in Figure (12, a). It starts with the entire training dataset at the root node. Then it selects the feature and split point that best divides the data into purer groups based on the target variable. Furthermore, this process repeated recursively on the divided subgroups, splitting them further based on additional features until they cannot be split more or reach a minimum size. Thereby, when making predictions, new data is sorted through the tree by going down branches based on its feature values until it reaches a final subgroup/leaf node. Consequently, the prediction is then the most common label of data in that subgroup. Therefore, this allows decision trees to represent complex patterns in data through an intuitive tree structure (Breiman, 2001, 2017; Quinlan, 1986).

However, single decision trees also exhibit some drawbacks. Specifically, they are prone to overfitting individual training examples, resulting in reduced generalizability to new data. Moreover, small variations in the training data can significantly impact the structure of the learned tree, diminishing stability. Additionally, decision trees often perform poorly when features include noise or exhibit low correlation with the target variable (Zhang, 2000).

One of the machine learning algorithms that construct the ensemble of multiple decision trees into a “forest” during training is random forest (RF) which shown in Figure (13, b). By averaging the prediction of these multiple trees’ prediction, the RF addressed decision tree limitation.

In this approach, each tree is created a bit differently than others. When splitting nodes in a tree, the algorithm only looks at a random subset of features instead of all features. Also, it only uses a sample of the full training data for each tree.

This randomness helps prevent the trees from just memorizing the training data. It makes the trees more independent and diverse. Once the forest is grown, predictions can be made. New data is put down each tree to get a prediction. The trees "vote" for the most common class (in

classification) or average the predictions (in regression). The forest prediction is the most popular vote among all the trees. Because we average multiple simpler models, Random Forests are generally more accurate than a single decision tree. They also reduce the risk of overfitting the training data (Breiman, 2001; Liaw & Wiener, 2002).

Random Forests work well for large complex datasets with many types of features. They are widely used today because they are accurate, easy to use, and can handle unbalanced or missing data. The randomization makes them robust models for both classification and regression problems (Genuer et al., 2010).

For an individual decision tree, the depth is an important parameter that affect the performance of RF, it represents the maximum extent possible until reach the terminal nodes (leaves). In decision tree when the depth increased it leads to overfitting. However, in RF each tree is

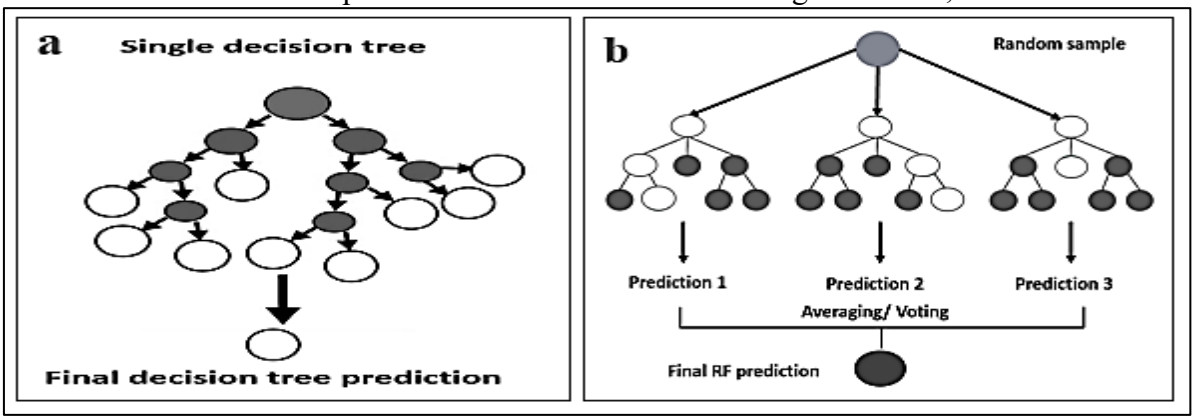


Figure 12: schematic illustration of decision tree (a), and random forest ML classifiers.

allowed to grow to the maximum extent possible until all terminal nodes contain less than a minimum number of training samples, typically set between 5-10. This approach prevents overfitting to small subsets of the data. While limiting tree depth can help control overfitting, random forests are less susceptible due to averaging effects between trees. However, deeper trees may result in better performance, up to a certain point before overfitting occurs. The optimal tree depth depends on the problem, and is usually selected through cross-validation during model tuning, with common values ranging from 5 to 15. Additionally, the effective depth of trees in a random forest is often lower on average than single trees due to random feature selection at each node (Breiman, 2001; Probst et al., 2019).

In this study, the extracted features from DenseNet-121 which were 1024 features/ image were used to train RF classifier for categorizing the dataset into five classes normal, hyper-acute, acute, sub-acute, and chronic. The RF consisted of 600 decision trees to balance accuracy with training time. Each tree was allowed to grow to an estimated maximum depth of 15-20 leaves, as is common for medical imaging datasets of this size and complexity. The trees were constructed through bootstrap sampling of the training features and random selection of a subset of features at each node split, helping to decorrelate the trees for more robust ensemble predictions. Cross-validation was employed to optimize hyperparameters and avoid overfitting. This two-stage approach leveraged transfer learning from a large-scale image database as well as the strengths of ensemble methods for the medical diagnosis task.

3.*4.2* Support Vector Machines (SVM)

Support vector machines (SVMs) are a powerful supervised machine learning algorithm for classification and regression that has gained significant attention in recent decades. Originally proposed by Cortes and Vapnik, SVMs function by finding an optimal hyperplane that maximizes the margin between classes of data in a transformed feature space. This hyperplane is determined based on the "support vectors", which are the training examples closest to the decision boundary (Cortes & Vapnik, 1995).

SVMs were later extended to perform non-linear classification through the use of "kernel functions", which implicitly map inputs to high-dimensional feature spaces via the "kernel trick" proposed by Aizerman (Aizerman, 1964). Common kernel functions employed in practice include linear, polynomial, radial basis function (RBF), and sigmoid kernels. This kernel-induced feature mapping technique allows SVMs to identify linear relationships between features that were previously nonlinear in the original input space. As a result of their ability to efficiently perform both linear and nonlinear classification/ regression, SVMs have been widely applied across diverse domains involving complex, high-dimensional datasets (Schölkopf & Smola, 2002). Their success is largely attributed to the use of kernel functions, which enable SVMs to learn complex decision boundaries by transforming the input space into higher-dimensional feature spaces where data may be more easily separated (Vapnik, 1999).

So, SVMs are a powerful supervised learning approach due to their capacity for flexible model fitting through kernelization, which maps inputs to enriched feature spaces where optimal hyperplane estimation can be achieved.

In this research, SVM classifier employed to analyze the DenseNet 121 extracted features from the different CT categories represented in the dataset. As described previously, SVMs are a powerful supervised machine learning algorithm that finds an optimal separating hyperplane to maximize margin between classes of data. By mapping our features into a higher dimensional space using the kernel trick, SVMs can learn complex non-linear decision boundaries that are well-suited for our classification task.

The SVM was trained on a subset of our CT scans where the category was known, allowing it to learn the patterns in radiomic features that characterize each class. It was then tested on a held-out set of scans to evaluate its ability to automatically categorize new cases as representing normal tissue or one of the strokes timepoints based solely on the extracted quantitative imaging features. This SVM classification approach provides an objective and data-driven means of analyzing subtle differences in brain CTs related to stroke at various stages.

3.*4*.3 Logistic Regression (LR)

Logistic regression is a generalized linear model used when the dependent variable is binary or dichotomous. Unlike linear regression which predicts outcomes on a continuous scale, logistic regression predicts the probability of an event occurring versus not occurring through modeling the logit transformation of the odds.

Specifically, the logistic regression model estimates the odds of an event occurring as the ratio of the probability an event occurs to the probability it does not occur. It models the logit, or the natural logarithm of these odds, as a linear combination of the predictor variables.

Mathematically, this takes the form:

$$logit(\pi) = ln(\pi/(1-\pi)) = \beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta kxk \quad (6)$$

Where π is the probability of an event occurring, 1-π is the probability of it not occurring, β0 is the intercept term, and β1…βk are the coefficients of the predictor variables x1…xk.

This logistic transformation allows modeling of a binary dependent variable while avoiding violations of assumptions like linearity and normality that occur in linear regression. The estimated coefficients can then be used to calculate odds ratios that indicate the change in odds of an event occurring with a one-unit increase in the corresponding predictor.

Logistic regression is widely applicable in fields involving categorical outcomes like the social and medical sciences. It provides a powerful tool for modeling dichotomous dependent variables and predicting probabilities of binary events (Hosmer Jr et al., 2013; Menard, 2002).

In this step of research method, and to perform logistic regression for classification, the data were arranged in a binary format by comparing each pathological category (hyper-acute, acute, sub-acute, chronic) against the normal category separately. This resulted in four separate binary classification tasks, with normal designated as the reference category.

The extracted radiomic features from DenseNet-121 which comprising the predictor variables, were then input into logistic regression models to classify each sample as either the reference normal category or one of the pathological categories. Logistic regression was chosen as the classification algorithm due its ability to model dichotomous outcomes and estimate the probability of class membership based on combinations of predictor variables (Hosmer Jr et al., 2013).

Through this methodology, the ability of the radiomic features to discriminate between normal tissue and each stage of cerebral infarction was evaluated using logistic regression models. This allowed the identification of imaging biomarkers with the potential to aid in the classification and characterization of different stages of ischemic CVA based on CT radiomic profiles.

3.5 Performance Evaluation Metrics

Evaluation metrics are crucial for accurately and objectively assessing the performance of machine learning classifiers. So, Proper evaluation of machine learning models requires the use of standardized metrics. Commonly reported metrics include accuracy, precision, recall, and F1 score, which provide insights into a model's abilities to correctly classify examples. However, these metrics can be influenced by data biases like class imbalance. The AUC metric calculated from ROC analysis evaluates classifiers across all classification thresholds in a manner independent of bias. Evaluation is also informed by measuring elapsed time, important for applications where efficiency is paramount. A comprehensive assessment of any proposed model utilizes each of these metrics to develop a thorough understanding of its predictive strengths, weaknesses, and computational feasibility for the given problem domain. The results under multiple evaluation criteria then guide further algorithm refinement and selection of the optimal approach (Powers, 2020).

To obtain a comprehensive assessment of model performance used in the research, using Python, metrics such as accuracy, precision, recall, F1 score, AUC, and elapsed time were calculated to characterize the predictive capabilities, bias, efficiency, and tradeoffs of different machine learning approaches for the given classification task.

3.5.1 Accuracy

Accuracy is a performance metric used to evaluate the effectiveness of a classification model. It measures the proportion of correct predictions made by the classifier over the total number of classification examples (Sokolova & Lapalme, 2009).

Mathematically, accuracy is defined as:

$$Accuracy = (TP + TN) / (P + N) \quad (7)$$

Where:

- TP represents the number of true positives, which is the number of examples correctly classified as positive.
- TN represents the number of true negatives, which is the number of examples correctly classified as negative.
- P is the total number of positive examples in the dataset.
- N is the total number of negative examples in the dataset.

By taking the sum of true positives (TP) and true negatives (TN), the total number of correct predictions were obtained. Dividing this by the total number of examples P + N gives the ratio of correctly predicted examples, expressed as a percentage between 0% and 100%. An accuracy of 100% indicates perfect classification with no errors, while lower values reflect more misclassified examples (Powers, 2020).

3.5.2 Precision

Precision is a classification performance metric that measures the ability of a classifier to identify only relevant instances (Sokolova & Lapalme, 2009). It is defined as the number of true positives (TP) divided by the total number of instances predicted as positive, including both true and false positives.

Mathematically, precision is expressed as:

$$Precision = TP / (TP + FP) \quad (8)$$

Where FP (false positives), is the number of instances incorrectly classified as positive when they are actually negative.

Precision ranges from 0 to 1, with a higher value indicating fewer incorrect positive predictions. A precision of 1 corresponds to perfect precision, meaning all positively predicted instances were indeed positive (Powers, 2020).

Precision alone does not provide a complete picture of classifier performance, so it is generally evaluated alongside other metrics such as recall and F1 score for a more comprehensive assessment (Sokolova & Lapalme, 2009). The precision-recall tradeoff also needs to be considered based on application requirements.

3.5.3 Recall

Recall in machine learning is a performance metric that measures the ability of a classifier to identify all positive instances. It is defined as the number of true positives (TP) divided by the total number of actual positive instances.

Mathematically, recall is expressed as:

$$Recall = TP / P \quad (9)$$

Recall ranges from 0 to 1, with a higher value indicating fewer actual positive instances are missed. A recall of 1 corresponds to perfect recall, meaning all positive instances are correctly identified.

Recall alone does not account for incorrect positive predictions, so it is generally evaluated alongside precision and other metrics for a complete analysis of classifier performance (Powers, 2020).

3.5.4 F1 score

The F1 score is a weighted harmonic mean of precision and recall, which provides an overall measure of a classifier's accuracy. It reaches its best value at 1 and worst at 0. The F1 score is calculated as:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (10)$$

The F1 score considers both precision and recall, which are two important and complementary metrics of classification performance. By calculating their harmonic mean, the F1 score

captures the tradeoff between precision and recall in a single number. A higher F1 score indicates better classification performance, with a maximum of 1 representing perfect precision and recall (Manning & Schutze, 1999).

3.5.5 Area under ROC curve (AUC)

The area under the receiver operating characteristic (ROC) curve, or AUC, is a metric used to assess the performance of a binary classifier. It ranges from 0 to 1, with higher values indicating better discriminative ability of the classifier.

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is also known as sensitivity, recall or probability of detection. The FPR is 1 – specificity or probability of false alarm.

The AUC can be interpreted as the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example (Fawcett, 2006).

The AUC is calculated as the area under the ROC curve. This can be estimated numerically using the trapezoidal rule as:

$$AUC = \frac{TPR1 + TPR2}{2} * (FPR2 - FPR1)..+(TPRn - 1 + TPRn)/2 * (FPRn - FPRn - 1) \quad (11)$$

Where TPRi, and FPRi represent the true positive rate and false positive rate at each threshold.

A higher AUC indicates better discriminative ability of the classifier to distinguish between positive and negative cases. An AUC of 0.5 indicates performance no better than random, while 1 represents a perfect classifier (Fawcett, 2006).

3.5.6 Elapsed time

Elapsed time is a metric used to evaluate the efficiency and computational performance of machine learning models, especially for applications where fast prediction speed is important. It measures the total time taken by a model to make predictions on a test dataset.

The elapsed time is usually calculated as the average time per prediction. It is obtained by:

- Recording the start time before making predictions on the tested data using the model.
- Recording the end time after all predictions are completed.

- Calculating the elapsed time as end time – start time.
- Dividing the total elapsed time by the number of samples in the test set to obtain the average time per prediction.

This is expressed mathematically as:

$$Elapsed\ time\ =\ (Te\ -\ Ts)\ /\ S \quad (12)$$

Where Te is the end prediction time, Ts is the starting prediction time and S the number of samples.

Units of elapsed time are typically milliseconds or seconds. Lower values indicate faster prediction speeds. Elapsed time evaluation helps identify models that meet the latency or throughput requirements of an application (Mohri et al., 2018). It is an important metric, especially for real-time systems or those processing large volumes of data.

3.6 Working Flowchart

The structural overview of the hybrid model is illustrated in Figure (14). This approach was specifically designed to support clinical stroke diagnosis, starting with dataset preparation and the acquisition of standardized images, to the pre-processing techniques that were used to improve the performance of the models, and to the features extraction step in which the output were 1024 radiomic features that finally used as input to train three classical machine learning classifiers. The SPEM image enhancement model played a crucial role in enhancing image clarity while maintaining the visibility of relevant structures and potential anomalies.

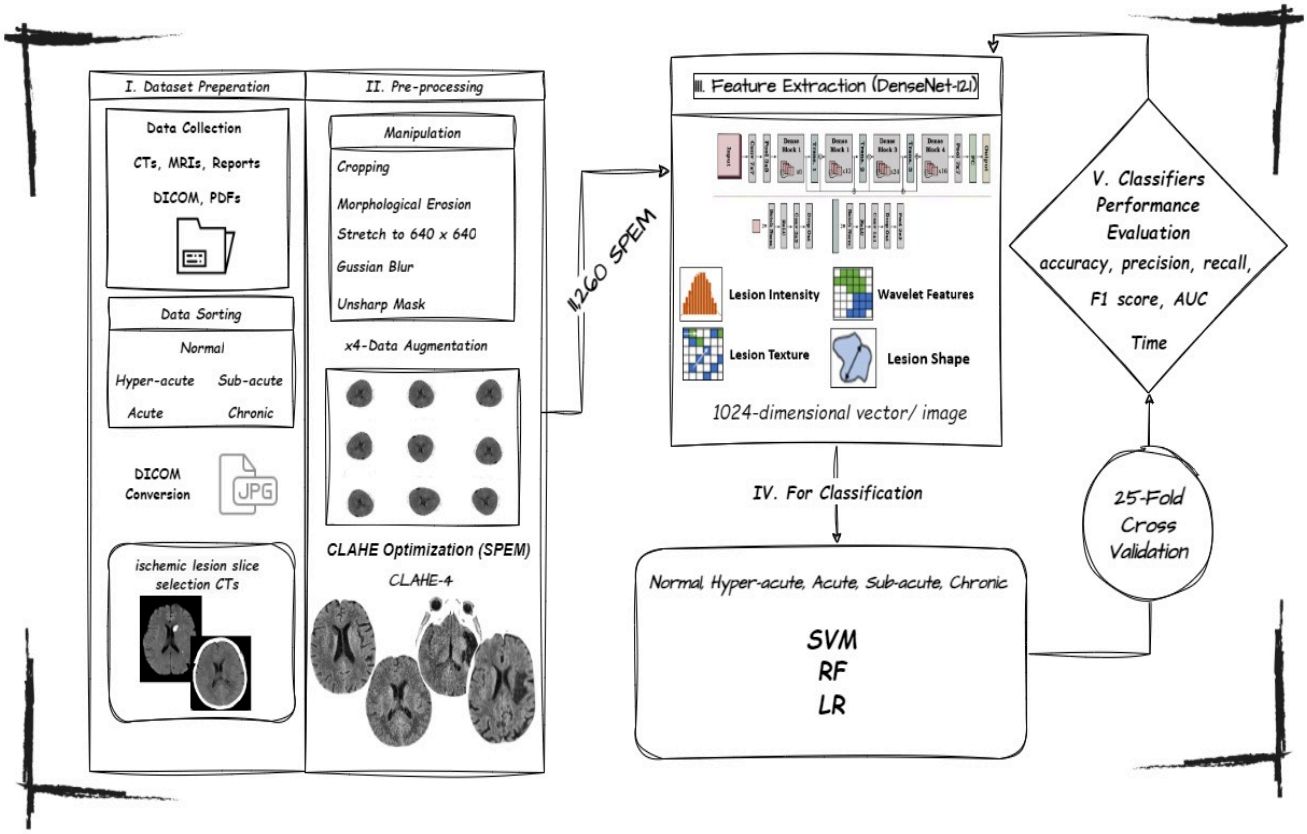


Figure 13: Architecture of the Hybrid Ensemble Deep Learning model. Created by Draw.io application.

Chapter four: Results

This chapter presents all research findings, these finding structured into three main stages. The first stage involved the evaluation of machine learning classifiers in classifying the extracted features by DenseNet-121 for B-NECT without pre-processing enhancement technique. The second stage including the evaluation of these classifiers' performance of the extracted features when the B-NECT images were enhanced using CLAHE at Clip Limit 4.

4.1 Optimal CLAHE Clip Limit Determination for SPEM Images

Table 2 displayed the values of the averages EME and PSNR of the preprocessed B-NECT. Noticeable improvement on image contrast observed at low levels, at Clip Limit 2 The EME had an average value of 16.115 while the PSNR was 21.399, indicating little change from the original images according to both metrics.

Table 2: Average EME and PSNR findings of B-NECT at different Clip Limits of CLAHE

Clip limit	2	2.5	3	4	8	20
Average EME	16.115	17.187	20.105	21.965*	20.056	17.199
Average PSNR	21.399	19.973	18.914	17.612*	15.952	14.946

As expected, higher values of Clip Limits led to greater contrast stretching but also began to produce unwanted artifacts. The EME dropped steadily, but the PSNR followed a downward trend, suggesting degradation of image quality. However, in levels 8 and 20, over-amplification of noise and artifacts noticed which led to image quality degrading, figure 14 represented the effect of different Clip Limites on CT image quality.

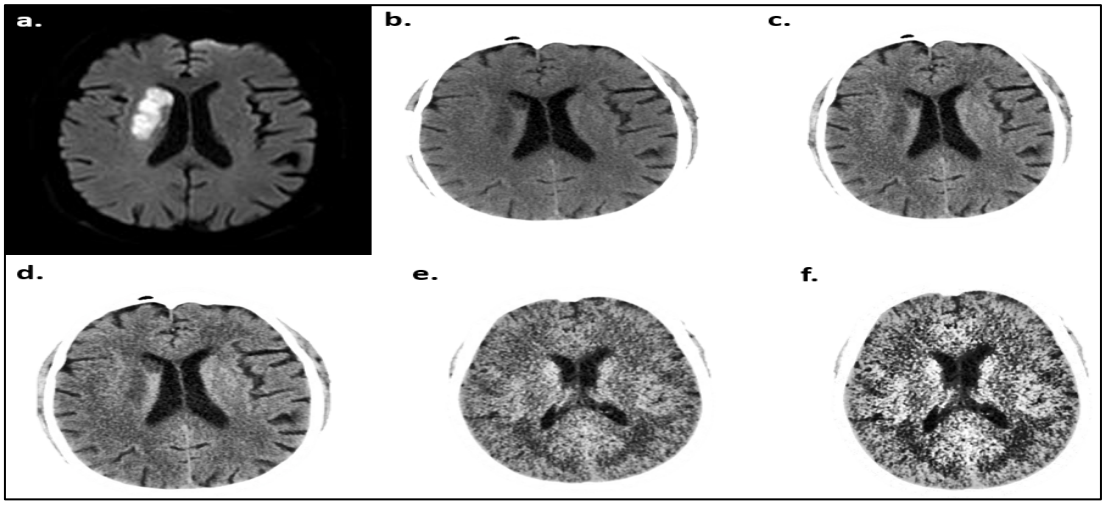


Figure 14: Acute ischemic stroke in the right caudate nucleus. (a) DW-MRI. (b) original B-NECT. (c- f) B-NECT with CLAHE at different Clip Limits (2, 4, 8, and 20) retrospectively

When the clip limit was at level 4, both EME and PSNR values saw an uptick. On inspection, the images exhibited strong yet natural-looking enhancement without visible noise. The average PSNR value was 17.612 and the EME was 21.965, indicating that the optimum balance point between enhancement and quality preservation was found. Figures 14 (d), 15, and 16 showed the resulted images when CLAHE- E was applied to B-NECT images across all ischemic stroke stages.

Through this systematic evaluation using objective metrics in Python version 3.9.18, the optimum balance point clip limit that produced the best results for the study dataset was determined to be 4 that optimized the images for the downstream feature extraction by DenseNet-121, and classification stages of this research.

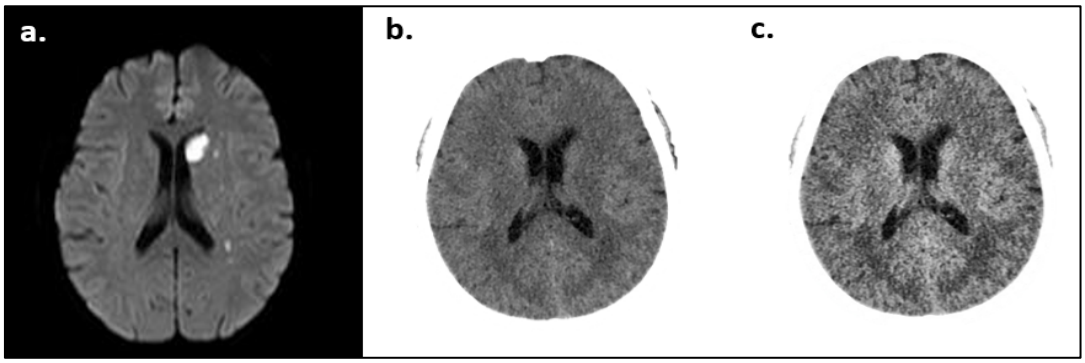


Figure 15: left caudate nucleus Hyper-acute ischemic stroke. (a) DW-MRI, (b) original B-NECT (c) B-NECT with CLAHE-4

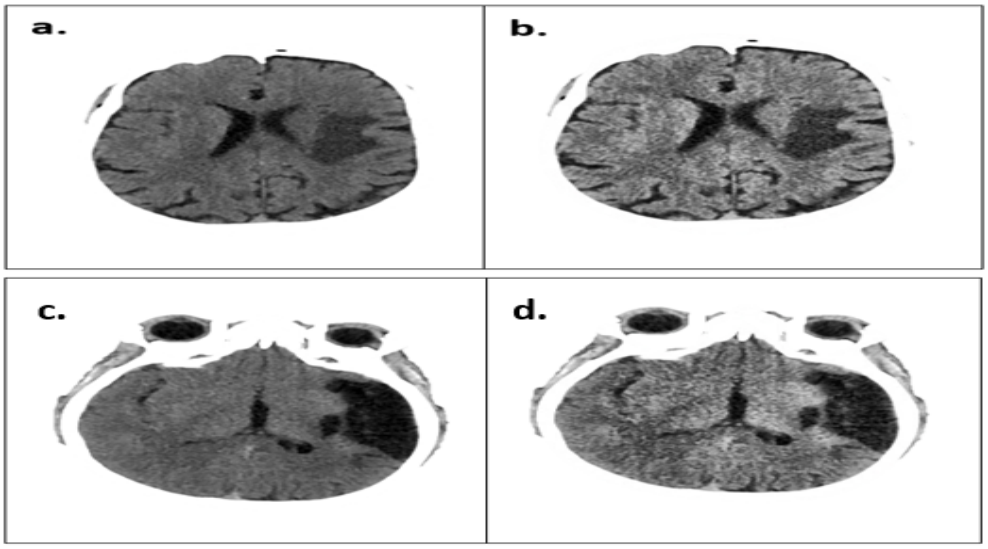


Figure 16: B-NECTs (a, b) showed fogging phenomena in the left cerebral hemisphere indicating sub-acute ischemic stroke (b) CLAHE-4. (c, d) B-NECT chronic ischemic lesion (c) original image, (d) CLAHE-4 image

4.2 Hybrid Model Performance Prior to Enhancement

4.2.1 Performance Comparison of the Model at Various Algorithms in the Hyper-acute Stage

Figure (17) shows a comparison of the models performance metrics in the hyper-acute ischemic stroke stage at different classifiers before the application of CLAHE-4.

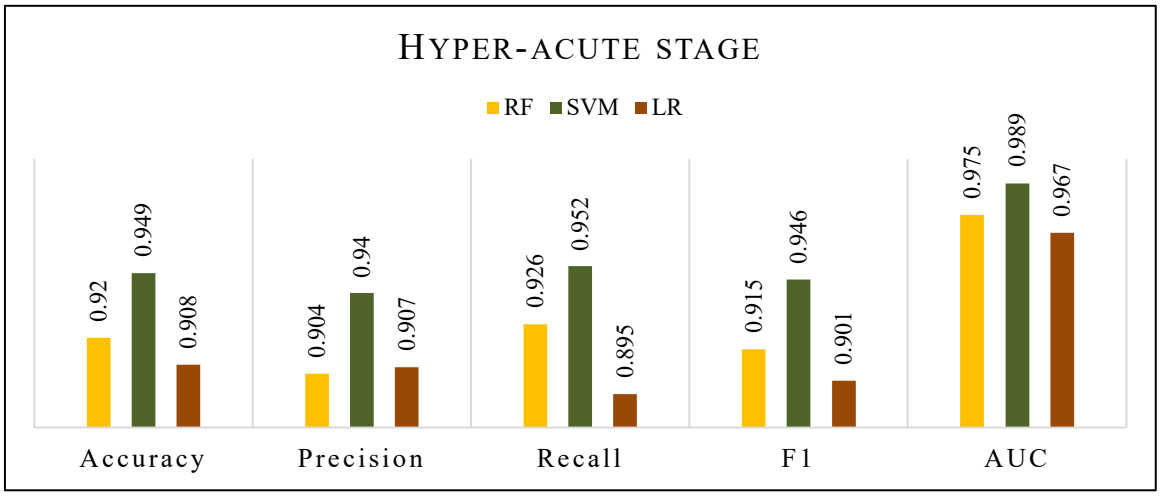


Figure 17: Hybrid models performance comparison in hyper-acute stage in non-enhanced datset.

For hyperacute ischemic strokes, the results showed that SVM achieved the highest classification accuracy of 94.9% with a precision of 94%, recall of 95.2%, and F1-score of 94.6% when applying the DenseNet121 features to the ischemic stroke lesion prediction task. SVM works by finding the optimal separating hyperplane between classes that maximizes the margin of separation. This allows it to discriminate strongly between positive and negative lesion cases, as evidenced by its highest AUC score of 98.9% out of all models.

RF obtained only a slightly lower accuracy than SVM of 92% while maintaining high precision of 90.4%, 92.6% recall, 91.5% F1-score, and 97.5% AUC comparable to SVM. This demonstrates that random forest was also highly effective at the classification. RF works by constructing many decision trees during training and outputting the class which is the mode of the individual trees' predictions. This tends to result in more stable performance compared to a single decision tree. In contrast, LR attained an accuracy of 90.8%, with 90.7% precision, 89.5% recall, 90.1% F1-score, and 96.7% AUC which is considered marginally lower than random forest.

However, on the prediction time side - presented in Figure (18) -, the most significant result was that LR was by far the fastest model, requiring a mere 0.4 seconds on average to make predictions. In contrast, the RF model achieved the second prediction speed with an average time of 41.87 seconds, while the SVM model had the longest elapsed time of 53.05 seconds.

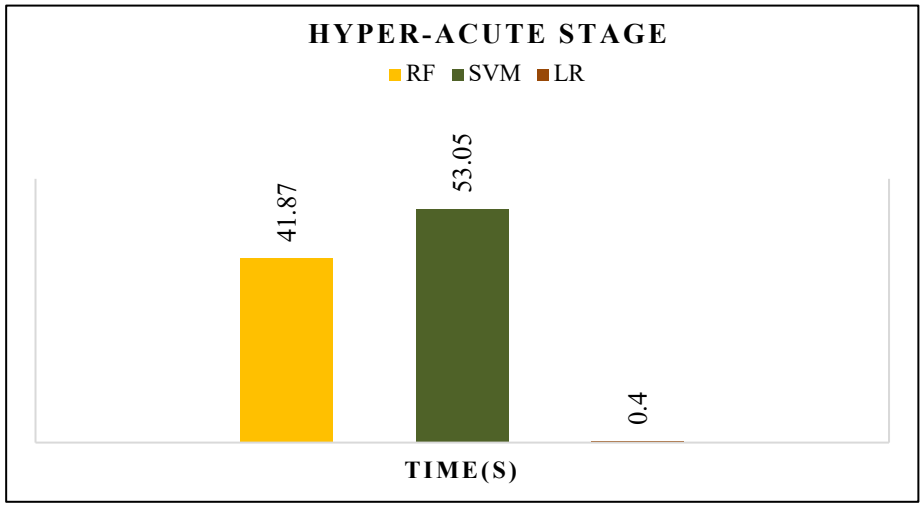


Figure 18: Comparison of hybrid models elapsed times in hyper-acute stage for non-enhanced dataset.

4.2.3 Performance Comparison of the Model at Various Algorithms in the Acute Stage

RF classifier in the acute stage, the model achieved an accuracy of 90.7% with a precision of 92.1%, recall of 92.7%, and F1 score of 92.4%. The AUC was 96.8%, demonstrating strong discriminative ability. For the SVM classifier, the model is achieved a higher accuracy of 93.9% in the acute stage, with a precision of 96.3%, recall of 93.7%, and F1 score of 95%. It also had a higher AUC of 98.7%, indicating excellent discrimination of lesions. LR attained the highest accuracy of 95.3% in the acute stage, matching SVM's precision of 96.3% and outperforming it on recall (96%) and F1 score (96.2%). LR also had a high AUC of 98.8%. Figure (19) shows a comparison of the models performance in the acute ischemic stroke stage at different classifiers before the application of CLAHE.

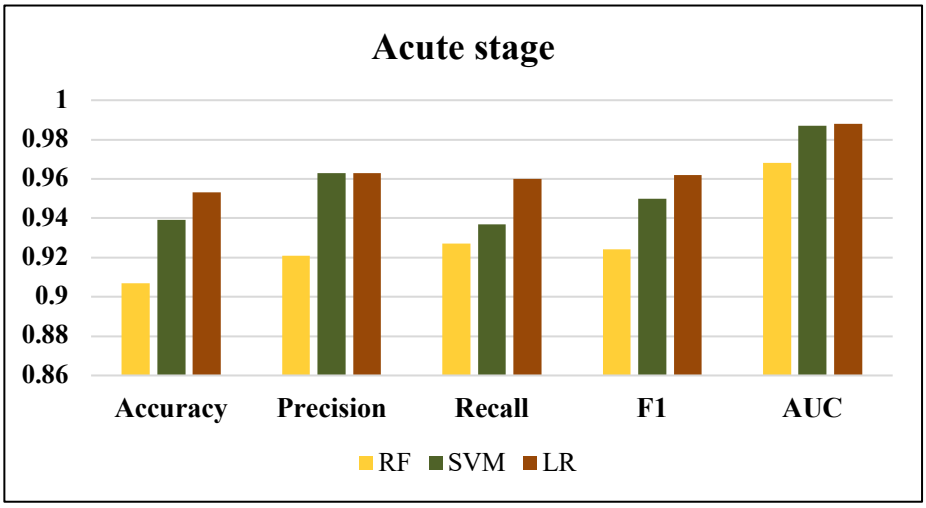


Figure 19: Hybrid models performance Comparison in acute stage for non-enhanced dataset.

While performance was good, RF was the slowest of the three models in the acute stage in which the average elapsed time for predictions was 8.63 seconds. SVM maintained very high performance levels but was not the fastest model with an average elapsed time of 2.52 seconds, making it faster than Random Forest but slower than LR. Most notably, LR was by far the fastest model with an elapsed time of only 0.06 seconds. While it matched SVM's performance, LR provided predictions over 40 times faster as shown in Figure (20). This makes it the best option for applications requiring extremely fast, real-time predictions in the acute stage.

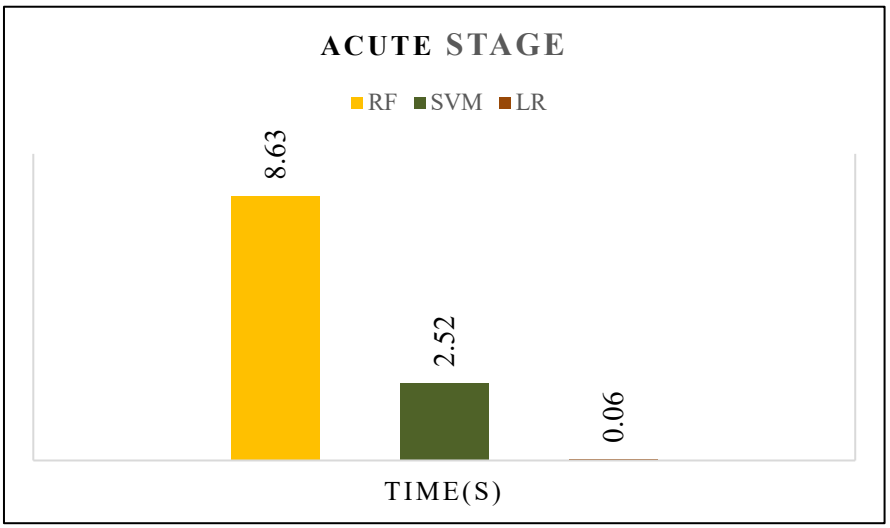


Figure 20: Comparison of hybrid models elapsed times in acute stage for the non-enhanced dataset.

4.2.3 Performance Comparison of the Model at Various Algorithms in the Sub-acute Stage

In the sub-acute stage, Figure (21) showed that SVM model achieved the highest accuracy of 96.3%, with precision and F1 scores of 98.6% and 95.8%. It also had a very high AUC of 99.2%. Moreover, SVM was faster than RF, requiring only 0.34 seconds for predictions. Nonetheless, its recall of 93.2% was lower than LR. Overall, SVM displayed best-in-class accuracy but was not quite the fastest model. In terms of accuracy LR model follows SVM with accuracy of 95.7% in the sub-acute stage. Furthermore, it had the highest recall of 96%. Consequently, LR matched or outperformed the other models on multiple metrics, Its precision, F1 score and AUC were excellent at 94.7%, 95.3% and 99.4% respectively. Howver, RF classifier achieved an accuracy of 93.3% with precision and recall scores of 93.2% and 91.9% respectively. Its F1 score was 92.5% and AUC was 99.2%, indicating very strong predictive performanceSVM attained the highest accuracy in the sub-acute stage of 96.3%, with precision and F1 scores of 98.6% and 95.8%. It also had a very high AUC of 99.2%.

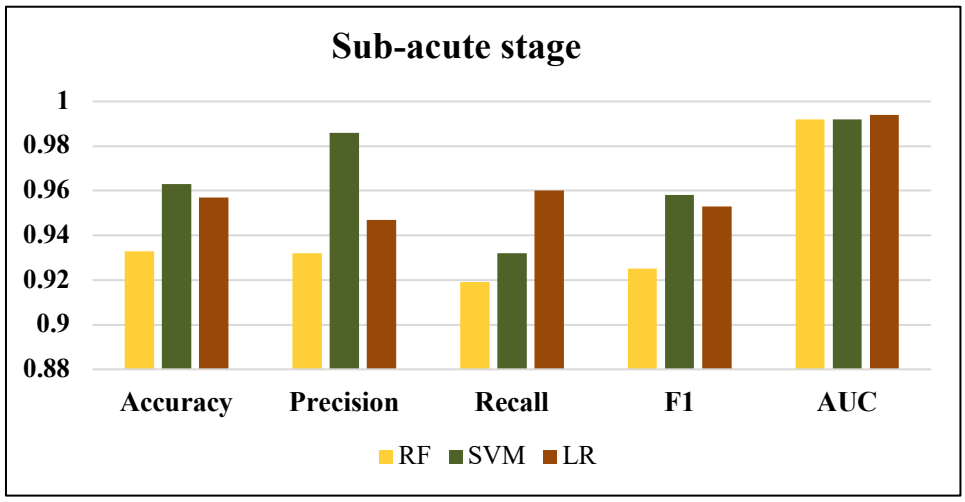


Figure 21: Hybrid models performance Comparison in sub-acute stage for non-enhanced dataset.

In the term of prediction speed illustrated in Figure (22), LR is considered the shortest elapsed time when it compared with SVM and RF models making it over 8 times faster than SVM and over 58 times faster than RF. It requested just 0.04s for prediction while RF model considered the slowest model with a time of 2.35s. Overall, SVM displayed best-in-class accuracy but was not quite the fastest model, it required 0.34s.

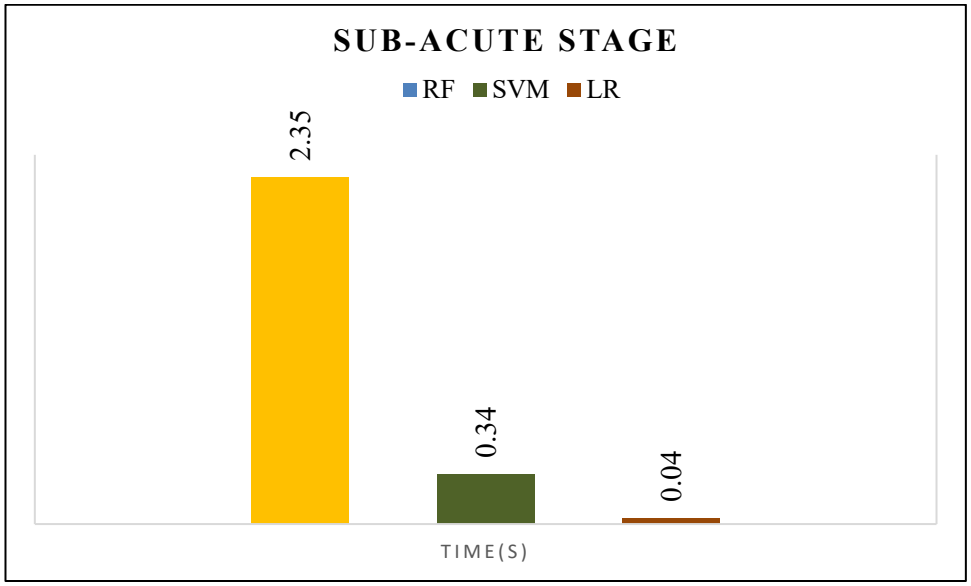


Figure 22: Comparison of hybrid models elapsed times in acute stage for the non-enhanced dataset.

4.2.4 Performance Comparison of the Model at Various Algorithms in the Chronic Stage

The results in Figure (23) showed a good classifiers' performance in the chronic stage. However, while RF achieved a respectable accuracy of 94.5%, precision of 95%, and recall of 94.3%, yielding a strong F1 score of 94.7% and AUC of 98.6%, its elapsed prediction time of 3.44 seconds was substantially slower than the other models as it clarified in the Figure (24). Therefore, though RF performed well, it was the least optimal choice for applications requiring rapid predictions.

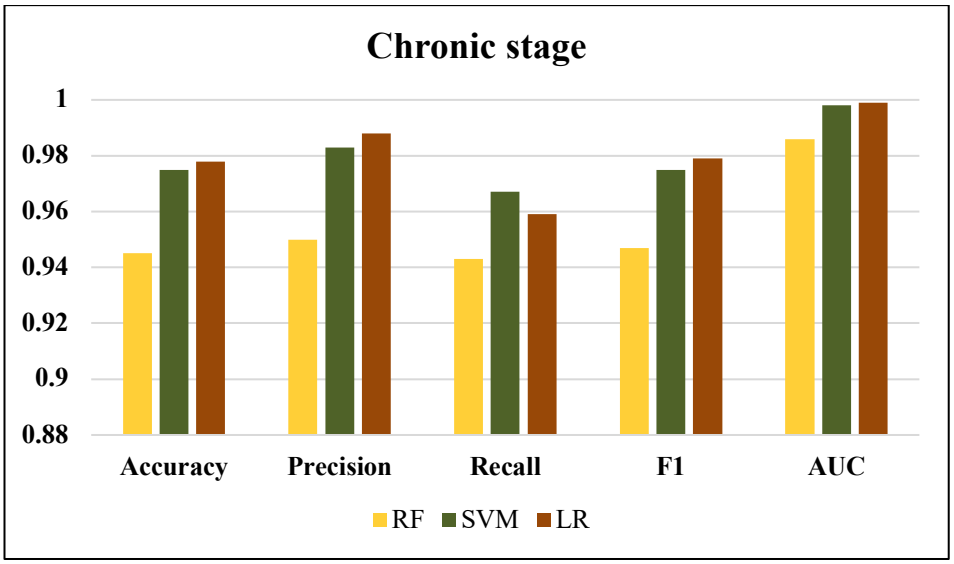


Figure 23: Hybrid models performance Comparison in chronic stage for non-enhanced dataset.

Meanwhile, SVM attained higher accuracy in the chronic stage at 97.5% and had an excellent F1 score of 97.5% and AUC of 99.8%. Moreover, its prediction speed of 0.61 seconds, while it was faster than RF, but was still slower than LR. Thus, while SVM displayed top-tier accuracy, it was not the fastest overall.

Most notably, LR achieved the highest accuracy of 97.8% along with the strongest precision of 98.8% and fastest prediction time of just 0.05 seconds. Furthermore, its recall of 95.9% and F1 score and AUC of 97.9% and 99.9% matched or exceeded the other models. Consequently, not only did LR demonstrate the best balance of high-performance metrics, but it was also over 10 times faster than RF and over 12 times faster than SVM. Therefore, for applications

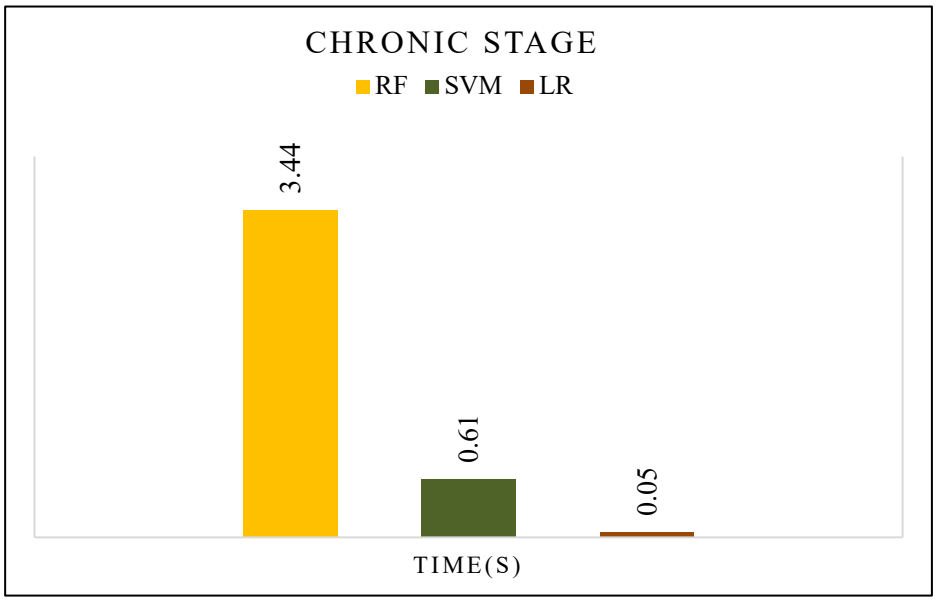


Figure 24: Hybrid models performance Comparison in chronic stage for non-enhanced dataset.

requiring both accuracy and real-time prediction speeds, LR proved to be the optimal classifier for the chronic stage.

4.3 Hybrid Model Performance with SPEM Image Enhancement

This section presents the results from experiments investigating the impact of CLAHE preprocessing on classifier performance for automated ischemic stroke classification from B-NECT images. Three machine learning algorithms – SVM, RF, and LR - were evaluated on their ability to identify stroke subtypes ranging from hyper-acute to chronic.

CLAHE with a clip limit of 4 was applied as a preprocessing step to enhance image intensities before feature extraction. Deep radiomic features were then extracted using DenseNet-121 from the SPEM images. These formed the input data used to train and test the classifiers.

The key focus of this section is to examine how CLAHE preprocessing with a clip limit of 4 before feature extraction influences subsequent classifier diagnostic ability, as assessed through performance metric comparisons across stroke subtypes and algorithms.

4.3.1 Performance Comparison of the Model at Various Algorithms in the Hyper-acute Stage

In the hyper-acute stage, while all classifiers performed exceptionally well, LR stood out as offering the best combination of accuracy, recall, and extremely rapid prediction times when

classifying hyper-acute lesions from SPEM images. Consequently, it is best suited for real-time applications.

In detail, all three classifiers achieved exceptional performance when applied to SPEM as displayed in the figures (25, and 26). RF attained very high accuracy of 98.32%, precision of 99.18%, F1 score of 98.38% and 99.95% AUC. However, its prediction time of 3.30s was the slowest. Therefore, while demonstrating strong predictive ability, RF was least optimal for applications requiring rapid results.

SVM achieved an accuracy of 99.15% with precision and recall scores of 98.30% and 99.45%. It also had an excellent F1 score and AUC. Furthermore, SVM was faster than RF, requiring only 0.52 seconds for predictions. However, it was still slower than LR. Thus, SVM displayed high performance but was not the fastest model.

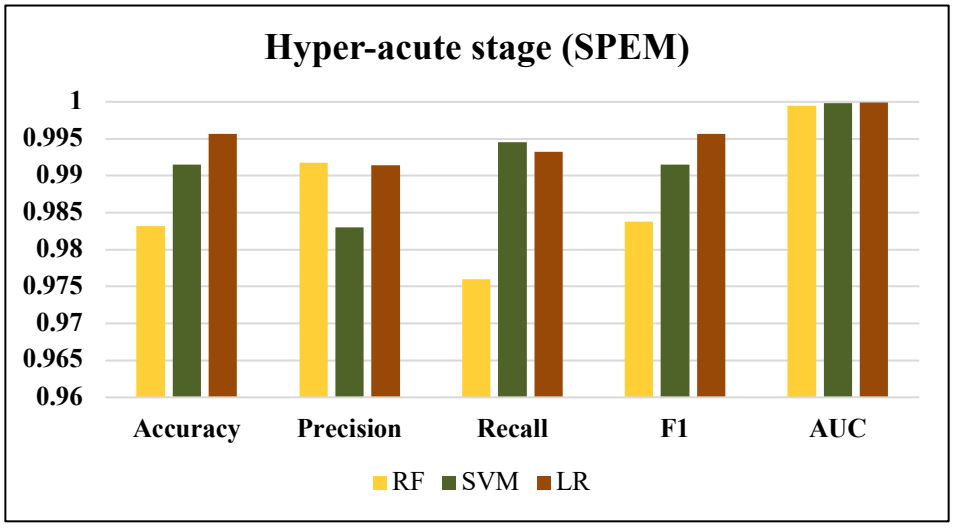


Figure 25: Hybrid models performance Comparison in hyper-acute stage for SPEM model.

LR attained the best performance with an accuracy of 99.57%, precision of 99.14%, recall of 99.32% and F1 score of 99.57%. Moreover, its AUC of 0.9999 demonstrated near-perfect discriminative ability. Most significantly, LR was the fastest model, requiring a mere 0.04

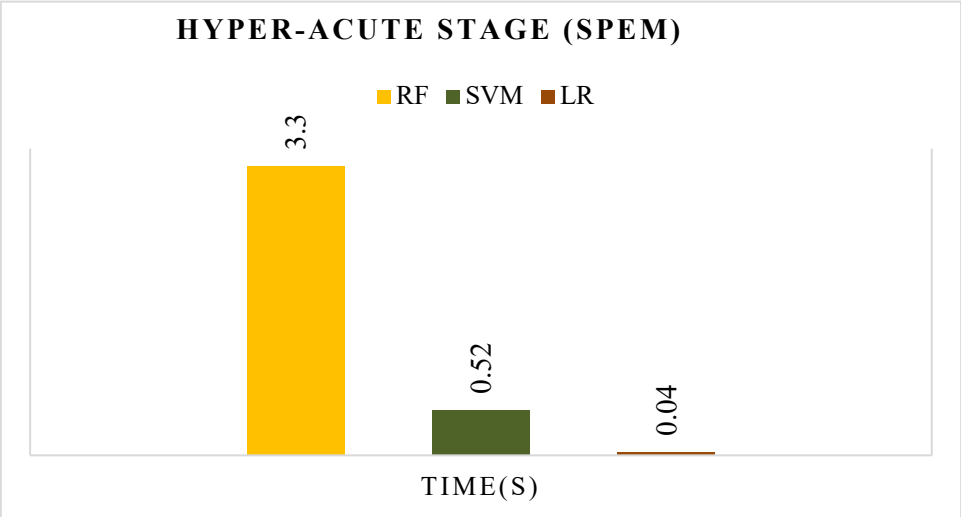


Figure 26: Comparison of hybrid models elapsed times in hyper-acute stage for SPEM model.

seconds for predictions. Consequently, not only did it achieve the highest accuracy, but it was over 100 times faster than the other classifiers. Therefore, LR proved to be the optimal choice for applications requiring both strong predictive capabilities and real-time prediction speeds.

4.3.2 Performance Comparison of the Model at Various Algorithms in the Acute Stage

In the acute stage after CLAHE application, all models performed well as illustrated in Figures 27 and 28, LR model achieved the highest accuracy while being over 100 times faster than the RF model and over 58 times than SVM, demonstrating it is the optimal classifier for real-time model applications. In general, SVM achieved the highest accuracy of 99.30% when applied to SPEM, with excellent precision, recall and F1 score of 99.31%, 99.30% and 99.30% respectively. It also demonstrated very strong discriminative ability with an AUC of 99.97%. However, its prediction time of 24.74 seconds, while faster than RF, was still much slower than LR.

As for LR, the model achieved an accuracy of 99.22% with precision, recall and F1 scores of 99.19%, matching SVM's performance. Impressively, it completed predictions within just 0.42 seconds, over 50 times faster than SVM and over 100 times faster than Random Forest.

On the other side, RF attained an accuracy of 97.16%, along with precision, recall and F1 scores of 95.34%, 98.96% and 97.11%. It also had a high AUC of 0.9961. Nonetheless, it was the slowest model, requiring 42.73 seconds for predictions.

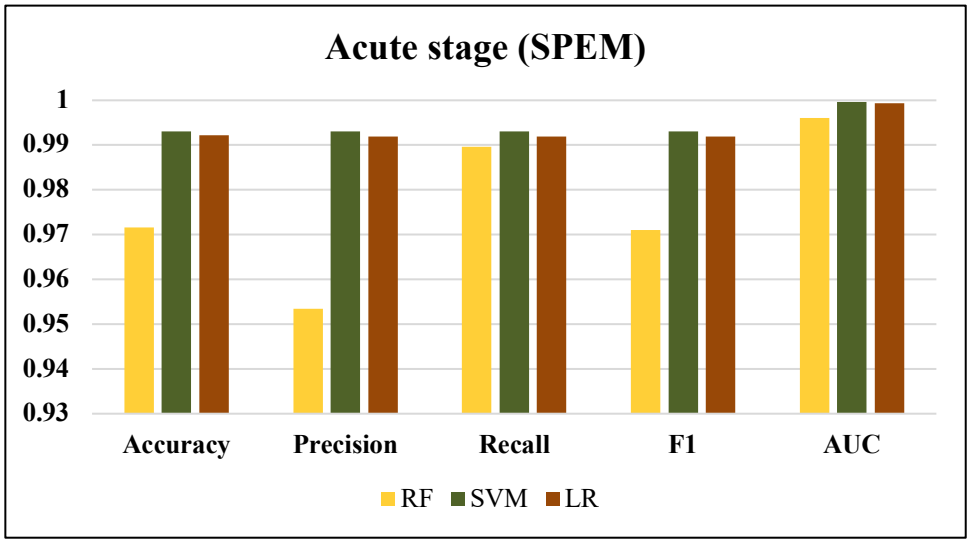


Figure 27: Hybrid models performance Comparison in acute stage for SPEM model.

Therefore, while SVM and LR demonstrated comparable high performance, LR stood out as the optimal choice for acute lesion classification from contrast-enhanced SPEM images due to its unparalleled prediction speed combined with excellent accuracy and other metrics. Its rapid results make it especially well-suited for time-critical applications.

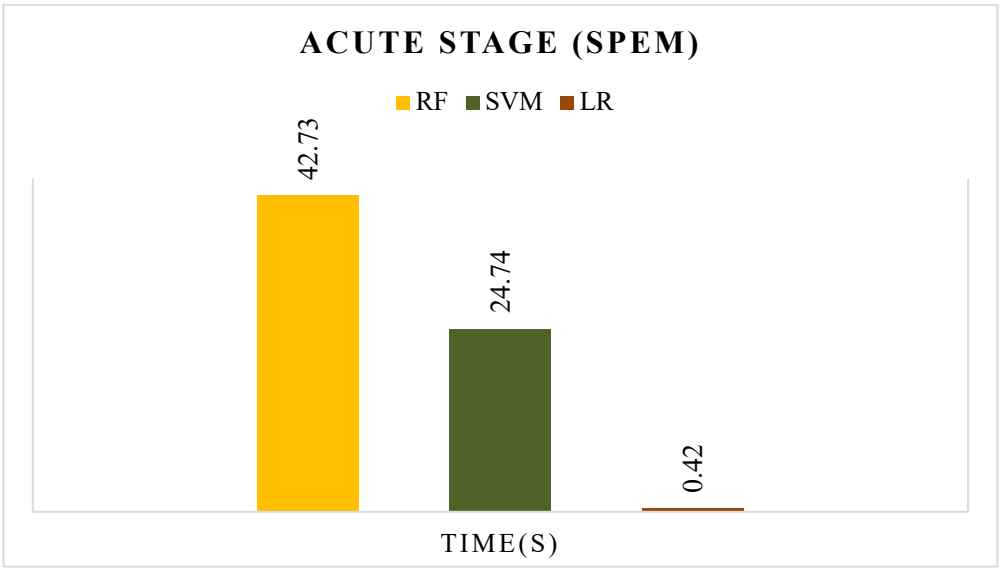


Figure 28: Comparison of hybrid models elapsed times in acute stage for SPEM model.

4.3.3 Performance Comparison of the Model at Various Algorithms in the Sub-acute Stage

In the sub-acute stage the performance results illustrated in the Figure (29), all models performed very well; however, LR model balanced strong accuracy with the fastest prediction speeds, demonstrating its effectiveness for real-time use cases. Specifically, as shown in Figure(30) LR model achieved the highest performance with an accuracy of 97.55% for classifying lesions from SPEM images. It demonstrated excellent precision, recall and F1 score of 97.55%, 97.53%, 97.53% and 97.53% respectively. Impressively, it completed predictions within a mere 0.03 seconds, over 10 times faster than SVM. While SVM matched LR's performance closely with an accuracy, precision, recall and F1 score of 96.39%, 98.76% and 97.56% respectively. SVM's discriminative ability was also very strong, with an AUC of 99.93%. Notably, it was faster than RF, requiring just 0.31 seconds for predictions.

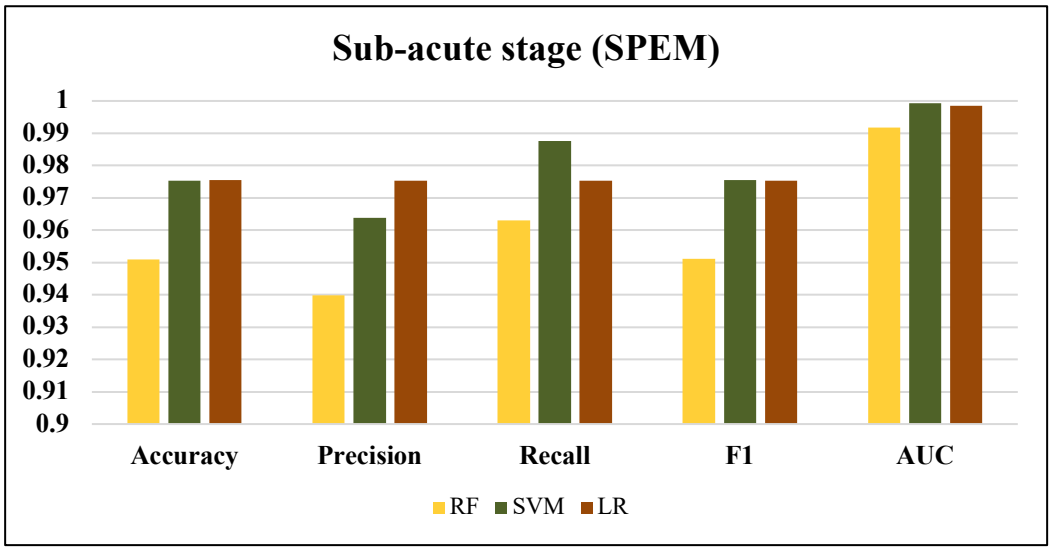


Figure 29: Hybrid models performance Comparison in sub-acute stage for SPEM model.

However, for RF model its attained an accuracy of 95.09% along with precision, recall and F1 scores of 93.98%, 96.30% and 95.12%. Its AUC was also high at 99.18%. However, it took the longest time of 2.07 seconds to make predictions.

While SVM and LR achieved comparable high accuracy and other metrics, LR model stood out as the preferred model for sub-acute lesion classification due to its unparalleled speed of

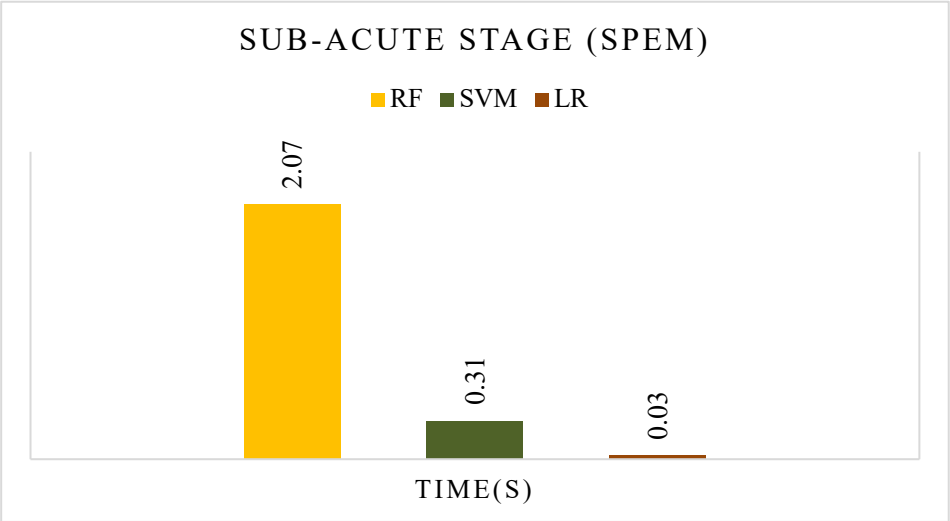


Figure 30: Comparison of hybrid models elapsed times in sub-acute stage for SPEM model.

just 0.03 seconds per prediction, combined with its excellent performance. Its rapid results indicate it is optimally suited for time-sensitive use cases.

4.3.4 Performance Comparison of the Model at Various Algorithms in the Chronic Stage

Generally, as shown in Figures 31 and 32, in the chronic stage all models performed well, but LR model stood out as offering the best combination of predictive accuracy, recall, and extremely rapid prediction times, making it particularly suitable for almost real-time applications involving chronic lesions.

More clearly, SVM achieved the highest accuracy of 99.59% for classifying lesions from SPEM images. It demonstrated excellent precision, recall and F1 score of 99.28%, 99.41% and 99.64% respectively. SVM's discriminative ability was nearly perfect, with an AUC of 99.99%. While it was faster than RF at 1.83s, it was still slower than LR. Despite of LR model matched SVM's performance closely with an accuracy, precision, recall and F1 score of 99.59%, 99.28%, 99.76% and 99.64%, it completed predictions within just 0.06s, over 30 times faster than SVM and over 100 times faster than RF.

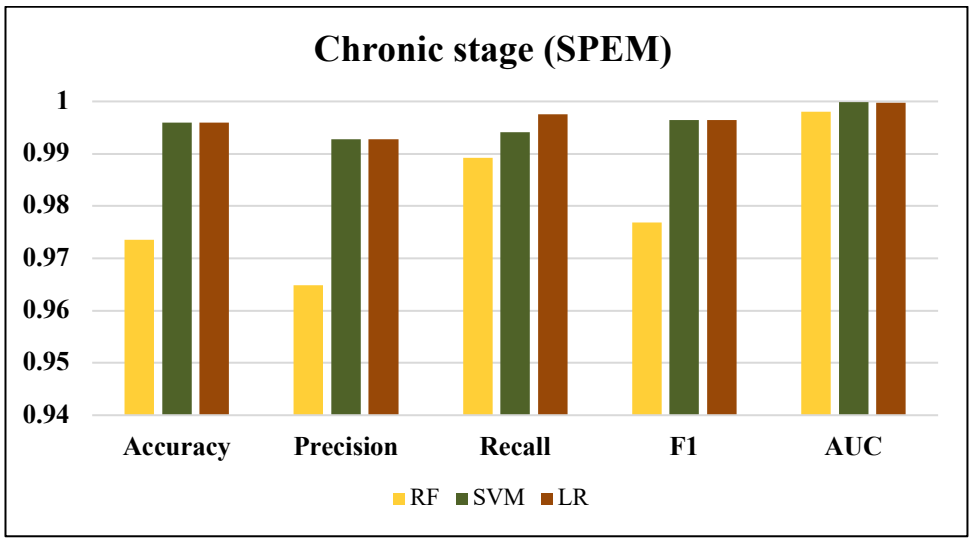


Figure 31: Hybrid models performance Comparison in chronic stage for SPEM model.

In contrast, RF model attained an accuracy of 97.36% along with precision, recall and F1 scores of 96.49%, 98.92% and 97.69%. Its AUC was also very high at 99.8%. However, it took the longest time of 8.36s to make predictions. Therefore, while SVM and LR demonstrated equivalent high predictive performance, Logistic Regression stood out as the preferred model for chronic lesion classification due to its unparalleled speed of 0.06 seconds per prediction combined with its excellent metrics. Its rapid results indicate it is optimally suited for time-critical chronic lesion assessment.

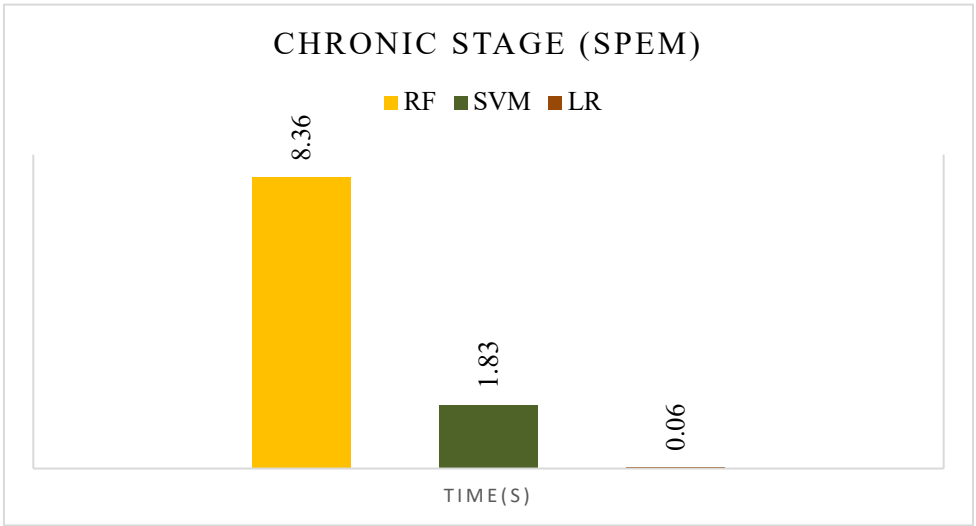


Figure 32: Comparison of hybrid models elapsed times in chronic stage for SPEM model

Chapter five: Discussion

Based in the proposed framework, this chapter discusses the construction of hybrid models and how was the experiment carried out. Also, the selection of optimum clip limit for CLAHE in the preprocessing step. Moreover, it discusses the realization of the architecture that presented in chapter three for the selection and evaluation of the three machine learning classifiers (RF, SVM and LR) used in building of the models.

5.1 Optimal CLAHE Clip Limit Determination for SPEM Images

CLAHE applied as pre-processing step to improve the image quality for the purpose of standardization of data representation to effectively enhance the performance of feature extraction from B-NECT images when it applied to DenseNet-121. Thus, lead to increase the performance efficiency of classifiers.

The clip limit parameter controls the contrast limiting and determines the amount of amplification applied across the image. The optimal clip limit parameter was determined through quantitative analysis at different clip limits (2, 2.5, 3, 4, 8 and 20) by measuring the EME and PSNR for the images. EME indicates the level of contrast enhancement achieved, with higher values representing better differentiation of intensities across the image. PSNR reflects image quality after processing, with higher dB values signifying less noise and distortion.

In this work, the results showed that a clip limit of 4 achieved the highest average EME of 21.965 while maintaining a good PSNR of 17.612 dB. Figure (33) describes the change that occurs to average EME and PSNR at different levels of clip limits (2, 2.5, 3, 4, 8, and 20). As seen, EME values generally increased from clip limits of 2 to 4, indicating improving contrast enhancement. However, from limits of 4 to 20, EME began to plateau and even decreased at higher limits (20). This suggests that amplification gains diminished or edge information lost beyond a limit of 4.

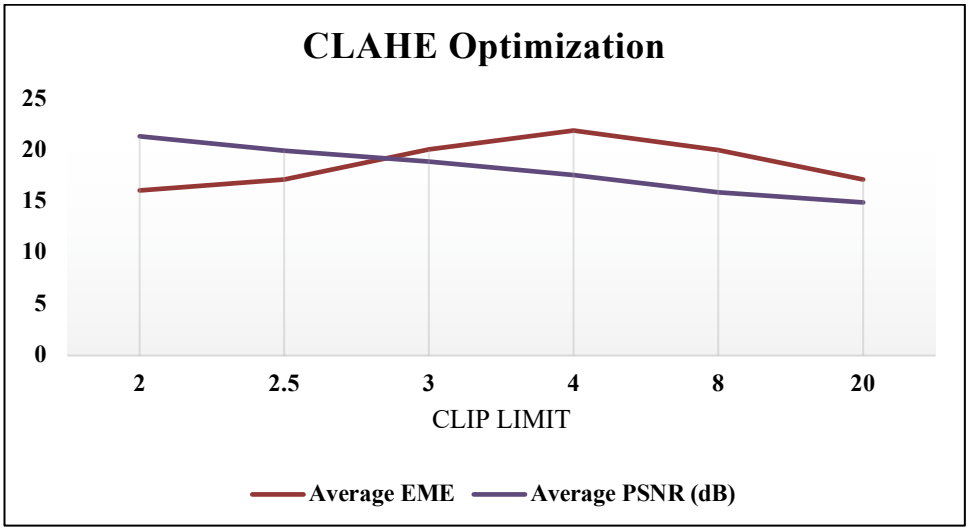


Figure 33: Chart illustrates CLAHE Clip Limit impact on both average EME and PSNR.

PSNR values remained high from limits of 2 to 3 but dropped steadily from 3 to 20. This drop was sharpest between limits of 4 to 8, and 8 to 20. This figure represents an optimal balance between maintaining image quality and preserving the intricacy of edges in the images. Our findings on image enhancement using the CLAHE algorithm align with prior studies, which have reported its effectiveness in enhancing image quality by preserving edge details and fine features. Furthermore, existing studies that used similar metrics for assessment, and which employed CLAHE for enhancing images, have reported a range of results that vary based on the characteristics and type of images utilized (Isa et al., 2017; Przelaskowski et al., 2007). The decreasing PSNR implies increasing noise and loss of useful detail in processed images at higher clip limits because a large Clip Limit value spreads out the histogram of grayscale pixels, amplifying noise but providing less contrast improvement. So, small Clip Limit value risks minimal enhancement which was described by Fawzi et al, to optimize the image enhancement (Fawzi et al., 2021).

By balancing the highest EME of 21.965 at limit 4 with an acceptable PSNR of 17.612, a clip limit of 4 provided the optimal trade-off between enhancement and preservation of quality for feature extraction. Therefore, based on quantitative analysis of evaluation metrics on a dataset of brain CT images, a CLAHE clip limit of 4 was selected as the parameterization that maximized pre-processing benefits for this study. The enhanced images with improved contrast aided in more accurate classification.

5.2 Hybrid Model Performance Before Enhancement

5.2.1 Performance Comparison of the Model at Various Algorithms in the Hyper-acute Stage

The study evaluated the performance of three different classifiers (RF, SVM, and LR) on their ability to classify five categories (normal, hyper-acute, acute, sub-acute, and chronic) on B-NECT features extracted by DenseNet-121 before image enhancement by CLAHE.

A similar trend is seen in the other evaluation metrics, with SVM achieving the highest precision, recall, and F1 scores out of the three classifiers tested. This indicates SVM makes the fewest mistakes in both correctly predicting and detecting true hyper-acute cases, achieving the best-balanced performance (García-Díaz et al., 2020).

The superior performance of SVM can likely be attributed to its maximal margin approach, which is well-suited for complex classification tasks involving high-dimensional feature spaces that contain more than 2000 dimensions (vectors) like those extracted by DenseNet-121. This high-dimensional nature poses challenges for other classifiers which are prone to overfitting such as logistic regression. However, SVM's intrinsic regularization properties derived from its maximal margin formulation allow it to learn robust decision boundaries even in very large dimensional spaces that conflict with the preventing of overfitting. As a result, this can indeed contribute to increased training time complexity compared to other classifiers compared with RF and LR retrospectively (53.05 s Vs. 41.87 s Vs. 0.4 s) (Statnikov et al., 2008).

RF classifier achieved excellent performance that was on par with the top-performing SVM model according to the evaluation metrics reported. Specifically, the random forest attained an accuracy of 92% and AUC of 97.5%, demonstrating a strong ability to discriminate between the normal and hyper-acute subtypes.

These results were slightly better than the logistic regression, which achieved an accuracy of 90.8% without other metrics provided for direct comparison. The RF's use of an ensemble of decision trees likely allowed it to model complex nonlinear relationships between the high-dimensional radiomic features extracted from the CT images better than the single decision boundary approach of logistic regression. Additionally, random forests tend to be more robust to outliers and are less likely to overfit than logistic regression given their averaging methodology.

These properties enabled the RF to outperform LR for this challenging classification problem involving subtle differences between stroke types in medical images. These findings were agreed with Belgiu and Dragut who discussed how RF average multiple decision trees to obtain better predictive performance than a single model as LR, with less variance and better handling of outliers (Belgiu & Drăguţ, 2016).

Breiman and Statnikov reported in their research that the fact beyond the shortest time shown in LR training is that it fits a simple linear decision boundary to the classification problem compared to other models. Logistic regression involves solving a convex optimization problem to determine the optimal weights, which can be efficiently approximated using gradient descent algorithms. This linear boundary is computationally inexpensive to calculate for large datasets compared to more complex nonlinear models. In contrast, RF and SVM models, have complex nonlinear relationships between features through an ensemble of decision trees and the determination of a maximal margin hyperplane, respectively.

These approaches provide superior performance but require solving non-convex optimization problems that are computationally intensive. This difference in computational complexity becomes greatly magnified during model training for high-dimensional medical imaging data with thousands of features (Bishop, 2006; Kim et al., 2005).

In this study on classifying CT images of stroke subtypes, logistic regression trained over 40 times faster than random forests, completing in just 0.4 seconds versus 41.87 seconds. Even faster support vector machines took around 53 seconds to train. While logistic regression performance was marginally lower, the ability to train a model nearly instantly provides advantages for time-critical applications like acute stroke diagnosis where rapid clinical decision-making is essential. Therefore, the shortest training time of logistic regression results from its computationally simple linear boundary formulation.

5.2.2 Performance Comparison of the Model at Various Algorithms in the Acute Stage

The models in this study achieved higher performance when classifying acute stroke cases compared to the hyper-acute stage on the CT image dataset, as evidenced by the increased accuracy, precision, recall, F1-score and AUC metrics. Specifically, the SVM model attained an accuracy of 93.9%, precision of 96.3%, recall of 96.3%, and F1-score and AUC of 97.2% for the acute stage, surpassing the metrics achieved for hyper-acute classification.

Additionally, the LR model classified acute strokes with a faster processing time of only 0.06 seconds compared to 8.63 seconds for RF in the hyper-acute stage. Faster prediction speeds are crucial for time-sensitive clinical workflows.

These results are consistent with the results of previous studies that have also observed increased classifier performance for later stage ischemic strokes that explained the reasons of this improvement. Previous work has also demonstrated higher reliability in acute stroke assessment scales versus hyper-acute scales. This suggests the definition of acute stroke may have been more consistent, presenting relatively "easier" classification cases, that also elucidated the general decreased time needed to classify the image in acute stage when compared with hyper-acute stage, which noticed in our results.

First, Peng and Chen were justified in their automated model that the visual features extracted by DenseNet-121 may have allowed for clearer separation between acute stroke and normal/other classes compared to hyper-acute. Moreover, DenseNet-121 has been shown to learn highly discriminative features for medical image classification tasks that improve with increased training data (Peng et al., 2022). He and Garcia were reviewed the challenges while training machine learning models, one of these challenges is overfitting negative impact that affect model performance. In this thesis, the acute stroke cases were potentially more numerous and balanced compared with hyper-acute stage (4000 acute Vs. 1260 hyper-acute images), which reduced the negative effects of overfitting known to hinder classifiers (He & Garcia, 2009).

In the case of RF performance and LR performance in both hyper-acute and acute stages, we can’t ignore that the number of datasets is 1260 images in hyper-acute while 4000 images in acute. This relatively lower images caused the LR have lower performance due to overfitting in the hyper-acute state resulted from the class imbalance. In fact, RF is better in the classification task due to the way of features analysis, in which the predictions resulted from all trees that employed it were averaged to overcome the overfitting problem, while LR performance increased with increased number of dataset as in acute stage (Liaw & Wiener, 2002).

On the other side, good classification performance was noticed in SVM in both hyper-acute and acute stages despite the difference of dataset size, this emphasizes the strong performance of SVM in limited data due to the kernelization process that enables hyperplane non-linear

high-dimensional feature space. SVM consistently achieved strong performance for both hyper-acute and acute stroke subtype classification, validating SVM as an effective algorithm for complex medical imaging problems involving thousands of discriminative features extracted from deep learning models, as also demonstrated in other studies as reported by Hebli et. al, in their study that used SVM for the classification purposes of brain tumor types using SVM classifier on brain MRI images (Hebli & Gupta, 2017). Moreover, the increased performance of SVMs and logistic regression over random forest implies the ability of these models to capture more complex nonlinear decision boundaries provided greater benefit for classifying acute strokes as reported by Abedi et, al. in their study (Abedi et al., 2020).

However, our findings showed that RF is slower when it compared with SVM in the case of acute stage which is contradicted hyper-acute result (8.36 s Vs. 2.52 s). To clarify the reasons, the algorithm nature of RF is constructed by ensemble of multiple independent decision trees (around 600 trees), when the task is for high-dimensional feature space, features were partitioned randomly and distributed to the trees in the 'forest'. For the individual tree to achieve the optimum decision RF must probe deeper. Additionally, the final output is the average of total prediction from all trees or the most common votes from all trees, so this process need longer prediction time when it compared with SVM where the algorithm nature utilize sparse solution and kernel trick application that relies a mapping between the input space and feature space for the approximations during the prediction in the high-dimensional feature space, that means the SVM need only to detect the signs to make decision (Genuer et al., 2010; Schölkopf & Smola, 2002).

5.2.3 Performance Comparison of the Model at Various Algorithms in the Sub-acute and Chronic Stages

The sub-acute stage classification performance using RF, SVM, and LR algorithms showed high accuracy. RF achieved an accuracy of 0.933, precision of 0.932, recall of 0.919, F1-score of 0.925, and AUC of 0.992, demonstrating strong discriminative ability to distinguish sub-acute cases.

SVM attained even higher accuracy of 0.963, precision of 0.986 indicating over 98% of positive predictions were true, recall of 0.932 correctly identifying 93.2% of actual positives, F1-score of 0.958, and AUC of 0.992 showing near-perfect classification.

LR yielded accuracy of 0.957, precision of 0.947 correctly classifying over 94.7% of cases, recall of 0.960 identifying 96% of true positives, F1-score of 0.953 and AUC of 0.994 both signifying excellent performance for sub-acute classification.

Regarding the chronic stage, the machine learning models demonstrated high performance based on CT imaging features. SVM achieved the highest accuracy of 97.5% for identifying chronic cases correctly. Notably, LR and SVM both attained near-perfect discrimination ability with AUCs exceeding 0.999, indicating exceptional ability to distinguish chronic from non-chronic cases. Precision scores were also extremely high for both LR and SVM at over 98%, suggesting a minimal rate of false positive predictions. While RF accuracy was slightly lower at 94.5%, its precision, recall, F1 and AUC metrics still signified strong classification capability for the chronic stage. Overall, the results confirm that chronic stroke manifestations have developed sufficiently clear visual patterns over time to be readily distinguished from other conditions or earlier stroke stages by sophisticated algorithms.

Upon reviewing the result, the models in sub-acute and chronic ischemic stroke stage showed improved performance for all classifiers, this is due to the increased ischemic stroke visual patterns on CT over time to later ischemic stages. Specifically, both the SVM and RF classifier exhibited accuracy gains of approximately 3-4% from the hyper-acute to chronic task. LR also improved by 2-5% between successive stages. This trend held for other evaluation metrics such as precision, recall and F1-score.

In general, the gradual enhancing of visual biomarkers over time thus appears to make later stage identification increasingly easier, benefiting all models as predictive performance rose in tandem with advancing pathology. Notably, the chronic classification yielded the peak results across SVM, RF and LR, corroborating it as the most discernible stage. These consistent improvements validate the potential for automated analysis to become more effective as stroke signs evolve on medical images.

Despite the increased visual patterns of ischemic stroke in CT over the time, the elapsed training times for the sub-acute classification task, in RF was 2.35 s, while SVM took 0.34 s, and LR 0.04 seconds. In contrast, for the chronic stage representing the most advanced pathology, RF time was 3.44 s, SVM 0.61 s, and LR 0.05 s, which mean that all classifiers in the chronic stage were slower than that in the sub-acute stage, Figure (34) illustrates the comparison between the prediction times for both stages. This suggests that as signs evolve the

furthest over the longest duration, the chronic classification problem imposed relatively more computational complexity for the models to solve compared to sub-acute. The trend of declining elapsed times from hyper-acute through acute and sub-acute stages validates that identification difficulty decreases as pathology progresses visually on imaging.

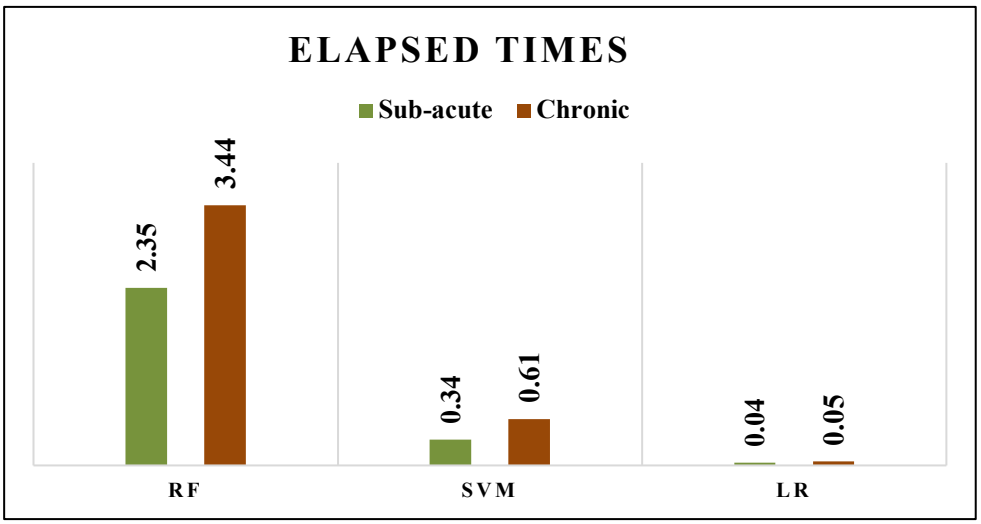


Figure 34: Comparison of hybrid models elapsed times in sub-acute and chronic stages for non-enhanced dataset.

5.3 Hybrid Model Performance with SPEM Image Enhancement

This discussion section analyzes results from experiments investigating the impact of CLAHE preprocessing on classifier performance for automated acute ischemic stroke classification from B-NECT images. Three machine learning algorithms – SVM, RF, and LR - were evaluated on their ability to identify stroke subtypes ranging from hyper-acute to chronic.

CLAHE with a clip limit of 4 was applied as a preprocessing step to enhance image intensities before feature extraction. Deep radiomic features were then extracted using DenseNet-121 from the SPEM images. These formed the input data used to train and test the classifiers.

The key focus of this analysis is to examine how CLAHE preprocessing with a clip limit of 4 prior to feature extraction influences subsequent classifier diagnostic ability, as assessed through performance metric comparisons across stroke subtypes and algorithms. Findings will also be discussed about previous work investigating similar preprocessing approaches.

Through investigating the impact of CLAHE, the aim is to provide insights that could help optimize automated acute stroke analysis methods. This may assist in developing more accurate clinical decision support tools derived from routinely acquired non-contrast CT scans.

5.3.1 Performance Comparison of the Model at Various Algorithms in the Hyper-acute Stage

The performance of machine learning models for automated classification of hyper-acute ischemic stroke using radiomic features extracted from B-NECT were evaluated in both with and without pre-processing image enhancement.

Findings provide valuable insights into optimizing machine learning approaches for hyper-acute ischemic stroke classification using CT imaging. The significant improvements observed across performance metrics when incorporating SPEM indicated that pre-processing plays an important role in feature extraction and subsequent classification.

By applying CLAHE with a clip limit of 4, SPEM enhances the contrast of subtle differences between healthy and pathological tissue, resulting in radiomic features that more clearly distinguish stroke subtypes. This allows classifiers to learn more discriminative patterns, improving their diagnostic ability.

The 5-9% absolute gains in accuracy, 3-5% increases in precision and recall, and 2-3% boosts to AUC seen with SPEM pre-processing demonstrate its ability to enhance subtle visual biomarkers that may otherwise be imperceptible without contrast adjustment. This agrees with prior work showing CLAHE improves lesion detection (Melingi et al., 2022).

Interestingly, training time was also substantially reduced, especially for SVM (53.05 s Vs. 0.52 s) and RF (41.87 s Vs. 3.30 s). This suggests SPEM normalization results in features with stronger class-separability, making classification an easier problem to solve computationally.

Notably, all three classifiers - including LR - benefited similarly from SPEM. This validates the pre-processing impacted the data itself rather than any particular learning algorithm.

By applying CLAHE to adjust the dynamic range and contrast of pixel intensities. This enhances subtle differences between healthy and pathological tissues in hyper-acute stroke CT scans, where lesions initially present with only slight contrast to surrounding brain parenchyma (Tan et al., 2012). By amplifying subtle intensity variations, SPEM enhancement model

extracted radiomic features from DenseNet-121 enable more clearly distinguishing of stroke subtypes based on their characteristic appearance patterns. Previous studies have shown CLAHE improves lesion detection sensitivity by enhancing low contrast regions (Tan et al., 2012).

Features extracted from enhanced images produce decision boundaries that are easier for classifiers to learn, resulting in more accurate classification. This agrees with Wang et al. (2020) who demonstrated data equalization enhancement techniques affects the transferability of deep neural networks by yielding features with stronger class-separability (X. Wang et al., 2020). So, incorporating SPEM enhancement prior to feature extraction using DenseNet-121 represents an effective strategy for optimizing hyper-acute stroke classification from non-contrast CTs. The consistent, significant gains across performance metrics validate its ability to enhance subtle visual patterns informative for diagnosis.

5.3.2 Performance Comparison of the Model at Various Algorithms in the Acute Stage

In acute stage, results showed significant performance improvement of the three classifiers - RF, SVM, and LR- when CLAHE-4 applied to CTs before the DenseNet-121 feature extraction process. Notably, when trained and tested on features extracted without preprocessing from raw images, RF achieved an accuracy of 90.7% while SVM and LR achieved 93.9% and 95.3% respectively. However, when using features extracted from SPEM-preprocessed images via DenseNet-121, all models demonstrated significantly improved classification metrics. RF accuracy increased to 97.2%, a 6.5% absolute improvement. Similarly, SVM accuracy improved by 5.5% to 99.3% and LR by 3.9% to 99.2%.

This improvement was also observed in other evaluation metrics as precision, recall and F1-score across all classifiers when using features extracted from SPEM preprocessed images via DenseNet-121 versus without preprocessing. This improvement indicated that the contrast adjustment technique plays an important role that optimized the model performance.

In RF, precision increased from 92.1% to 95.34% when SPEM features were trained which means false positive predictions reduced. Recall increasing from 92.7% to 98.96% indicating better identification of true positives (acute stroke) detected. Correspondingly, F1-score rose from 92.4% to 97.11% reflects an overall better balance between precision and recall achieved through SPEM. The F1-score considers both precision and recall, so its increase validates that both false positives and false negatives were reduced. Furthermore, the AUC relatively

increased when compared with features without CLAHE demonstrating that SPEM features allowed RF to achieve the near-perfect ability for acute ischemic stroke prediction, in other words, increased ability true positives from false positives.

On the other hand, significant improvement was also noticed in the performance of SVM when trained on features directly extracted from non-preprocessed images, SVM attained an accuracy of 93.9%, precision of 96.3%, and F1-score of 95%. In contrast, after applying CLAHE-4 preprocessing, SVM's accuracy increased to 99.3%, representing a 6% boost. Precision and F1-score also improved to 99.3%, gains of 3% and 4.5% respectively. Additionally, SVM achieved near-perfect recall and AUC metrics of 99.3% and 99.997%.In LR, SPEM model features added improvement to the accuracy compared with non-preprocessed features (99.2% Vs. 95.3%). Moreover, precision and recall both rose (96.3% Vs. 99.2%) and (96% Vs.99.2%) retrospectively, while AUC increased from 98.8% to 99.9%. These improvements indicate that CLAHE-4 enhancement amplified the subtle differences in radiodensity patterns in the acute prediction subtype. This allowed SVM, RF, and LR to learn more discriminative features, resulting in a more effective decision boundary with fewer misclassifications.

However, the elapsed time considered a critical consideration in medical applications especially in which real-time decisions must be as soon as possible as in acute and hyper-acute stroke stages. In this experimental assessment, the result showed that LR stands out as the most efficient computationally algorithm in the classification and prediction of acute ischemic stroke from high-dimensional extracted SPEM features from DenseNet-121. In fact, LR on average required less than a half second (0.42s) for the prediction task which considered extremely rapid, while SVM required nearly more than 59 times than LR (24.74s Vs. 0.42s), and RF required over 100 times longer than LR (42.73s Vs. 0.42s) made it the slowest classifier in acute stage prediction.

This extreme speed in LR is due to the computational simplicity to fit the logistic function for class probability estimation. Moreover, the linear relationship between decision boundaries that allowing very fast calculation of class prediction compared to non-linear algorithms like SVM and RF. Furthermore, the parallelization nature of LR; refers to the ability to split up computational tasks and perform them simultaneously across multiple processing units, such as the multiple cores in a CPU or the compute units in a GPU, this enables LR to deal with high-dimension and complex dataset (Hastie et al., 2009).

Concerning RF, this too long elapsed time is due to the algorithm's nature depending on the type of features, in this study the dataset was complex high-dimensional radiomic features led to increase the depth for individual decision tree and the need of large number of trees for accurate prediction outcome. While on the subject of SVM, the superior speed performance in classification is due to the computational complexity of the maximum hyperplane of classes by the use of kernel functions that considered computationally intensive during model training for high-dimensional medical imaging data with thousands of features (Bishop, 2006; Kim et al., 2005).

The prediction elapsed times for the hyper-acute and acute stroke classification tasks using non-preprocessed versus SPEM preprocessed features provided interesting insights. Without CLAHE, classifying hyper-acute cases required significantly longer prediction times across all classifiers relative to acute cases. For instance, SVM predictions on hyper-acute data took an average of 53.05s compared to only 2.52s for acute cases. However, when trained on features extracted from SPEM preprocessed images, this was reversed, the acute stroke stage necessitated markedly elevated prediction times for all models relative to hyper-acute. Using SPEM features, SVM prediction time increased to 24.74s for acute versus 0.52s for hyper-acute. These findings suggest that appropriate preprocessing like SPEM can substantially impact computational efficiency by altering the relative complexity of classification problems.

5.3.3 Performance Comparison of the Model at Various Algorithms in the Sub-acute Stage

In the sub-acute stage, the results demonstrated that applying CLAHE-4 in SPEM models led to enhanced classification and prediction across all three machine learning classifiers – SVM, RF, and LR- performance compared to when trained on features directly extracted from non-preprocessed CT images.

The result showed that the SPEM features extracted by DenseNet-121 were improved RF classification performance for predicting the sub-acute stroke stage. This is due to the filtering out of noise and irrelevant variation that is unrelated to the prediction task. For accuracy, increased from 93.3% to 95.09% with SPEM preprocessing, which suggested better classifying the sub-acute cases correctly. While precision increased slightly from 93.2% to 93.98%, which means fewer false positive predictions. Notable increase was noticed in recall (91.9% Vs. 96.3%) means the model is finding more of the true sub-acute stroke cases. The F1 score Increased from 92.5% to 95.12%. The F1-score considers both precision and recall and their

balanced harmonic mean. So, higher F1 score indicated better overall prediction performance considering both precision and recall. This means that SPEM enhances the model's balanced predictive ability. No significant change in AUC remaining it nearly 99.2% which reflect the high ability of the model to distinguish between sub-acute cases and non-cases of sub-acute.

SVM performance metrics showed improved prediction in hyper-planning separation between sub-acute cases and non-cases of sub-acute. Generally, improved performance was observed on SPEM preprocessed features when it compared with non pre-processed. Slightly increased of accuracy (96.3% Vs. 97.54%), F1-score (95.8% Vs. 97.56%) indicated better overall classification accuracy and balanced performance. More notably recall increased (93.2% Vs. 98.76%) demonstrating the SVM's enhanced ability to identify a higher percentage of true sub-acute cases. The near-perfect AUC for SPEM preprocessing (99.93% Vs. 99.2%) underscores the model's excellent discriminative power between classes.

Furthermore, the results show that SPEM preprocessing led to improved performance of the LR classifier for sub-acute stroke stage prediction. LR estimates the probabilities of different outcome classes by learning optimal coefficients for the features. All key evaluation metrics saw small gains with SPEM: accuracy and F1-score increased slightly, indicating better overall classification. Precision and recall increased more noticeably, demonstrating enhanced ability to precisely predict true positives and identify actual cases respectively. The near-perfect AUC of 99.86% for SPEM underscores the model's excellent discriminative power between classes.

General decreased elapsed time needed for the prediction of sub-acute SPEM feature in all classifiers when were compared with non-preprocessed features, due to the removal of irrelevant features that cause increasing in the computational time.

In summary, by filtering noise and retaining informative patterns using CLAHE-4, SPEM preprocessing provided RF, SVM, and LR with representation allowing to better learn the relationships between features and outcomes. This highlights the value of preprocessing for enhancing standard classifier performance.

5.3.4 Performance Comparison of the Model at Various Algorithms in the Chronic Stage

The results demonstrated that the SPEM pre-processed features extracted by DenseNet-121 from B-NECT chronic cases led to improved performance of the RF classifier for predicting chronic stroke stage.

Better classification accuracy observed (94.3% Vs. 97.36%) suggested increased ability of RF to true chronic ischemic stroke, Precision increased from 94.3% to 96.49%, indicating more reliable prediction of chronic cases. Recall substantially increased from 94.3% to 98.92%, demonstrating greatly enhanced detection of actual chronic patients. The F1-score rose slightly from 94.7% to 97.69%, reflecting overall improved performance. Notably, the AUC increased from 98.6% to 99.8%, representing much stronger discriminative power between classes. While training time took longer at 8.36 seconds compared to 3.44 previously, the sizeable improvements across key predictive metrics like precision, recall and AUC achieved through SPEM preprocessing demonstrate its value for providing RF more effective representation of the imaging features to learn from.

SVM saw even larger gains - accuracy improved more significantly from 97.5% to 99.59%, a 2.21% boost, outperforming non-preprocessed features. Precision increased more noticeably from 98.3% to 99.28%, reflecting more reliable positive predictions, while recall rose more substantially from 96.7% to 99.41%, demonstrating better sensitivity. F1-score likewise saw a larger increase from 97.5% to 99.64%, and AUC rose remarkably from 99.8% to near perfect 99.99%, underscoring vastly improved class separation ability with SPEM.

LR also benefited, with accuracy rising from 97.8% to 99.59%. Precision and recall both increased noticeably from 98.8% to 99.28% and 95.9% to 99.76% respectively, and F1-score rose similarly from 97.9% to 99.64%. AUC increased slightly from an already high 99.9% to 99.97%.

However, training time increased for all models when using SPEM preprocessed features, this was likely due to the additional computations required during the preprocessing stage to filter irrelevant variations and extract core patterns. Specifically, RF training time rose from 3.44 to 8.36 seconds. For SVM it increased from 0.61 to 1.83 seconds. And LR saw only a minor increase from 0.05 to 0.06 seconds. However, the substantial boosts to key predictive performance metrics like accuracy, precision, recall and AUC achieved through the enriched

feature representations provided by SPEM preprocessing far outweighed the marginal increases in training time. This suggests the value of incorporating SPEM as a data preprocessing step to enhance machine learning models for chronic stroke prediction applications.

5.4 Impact of DenseNet-121 and CLAHE in Models Performance Summery

5.4.1 DenseNet-121

The use of DenseNet-121 as a deep-CNN feature extractor strength the machine learning classifiers performance. compared with previous study if the application of CLAHE-4 as an image enhancement technique ignored. In this study, the results demonstrated a significant improvement when the extracted features by DenseNet-121 coupled with LR as a final classifier the 90% and 95% accuracy with high other metrics performance for hyper-acute and acute stage retrospectively compared with other studies. For example, Hyunna et al. in 2020, reached the greatest sensitivity of 75% and 82.6%, 83.1% specificity and AUC retrospectively when applied to features extracted from different image features descriptors or extractors as GLCM, gray level run-length matrices (GLRLM) from FLAIR and DWI-MRI images. Moreover, Gangavarapu and Gorli in 2021 reported in there analytical research of the performance of ML classifiers in stroke prediction that the LR reached accuracy and precision of 78% and 77.5% retrospectively, and 77.6% F1-score (Lee et al., 2020; Sailasya & Kumari, 2021).

In the other side, the performance of RF In this study showed significant performance improvement in classification of DenseNet-121 features, we can notice from performance evaluation that the RF accuracy was 92% with AUC of 97.5% for hyper-acute stage, and 90.7% with AUC of 96.8% in acute stage. Compared with previously Kniep et al. paper in 2020 who proposed RF to classify early ischemic changes in 20 most important B-NECT extracted features by PyRadiomics Python package. However, Hanning results focused solely on AUC which their result showed 82% for thalamus and 70% for cerebellum area.

While DenseNet-121 features classification using SVM showed high performance of 95% accuracy, 94% precision, and 99% AUC in hyper-acute stage and 94% accuracy, 96% precision, and 99% accuracy in acute stage compared with Roman et al, 2017 who evaluated

the SVM performance by measuring the AUC score of 94%-96% in classifying features extracted by GCM and RLM form registered B-NECT with DWI-MRI images. Also, Peixoto et al. in 2018, measured the specificity, sensitivity, F1-score, and accuracy of SVM to distinguish and classify B-NECT to healthy, or ischemic, or hemorrhagic stroke. He founded that 99%, 97%, 98%, and 98% retrospectively (Peixoto & Rebouças Filho, 2018; Peter et al., 2017; Roman Peter. Panagiotis Korfiatis, 2017).

In contrast to the previous study, our study undertook comprehensive evaluation of three classifiers performance including accuracy, precision, recall, F1-score and AUC. This extensive assessment increased the robustness of model building. Moreover, the time is a critical measurement especially in clinical application where the discesion must be as fast as possible for the essential treatment and management planning which the previous studies ignored it. Our study, however introduces a novel dimension by incorporating elapsed time as an essential metric, addressing an aspect that has been largely overlooked in the existing literature.

So, in summary, the use of DenseNet-121 for feature extraction likely contributed significantly to the strong performance achieved across classifiers. Some ways DenseNet-121 could have positively impacted the results:

- Representation power: DenseNet-121 is able to learn rich, hierarchical representations from raw pixel data due to its dense connectivity. This results in highly discriminative features that capture subtle visual patterns related to ischemic stroke subtypes.
- Generalization: DenseNet-121 is pretrained on large natural image datasets, providing a good starting point for the model to generalize to new medical image data. This transfer learning approach helps prevent overfitting.
- Feature reuse: Dense blocks allow earlier layer features to be reused extensively, improving information flow and regularization. This leads to more robust features compared to traditional architectures.
- Dimensionality: DenseNet-121 extracts thousands of features, providing classifiers with a very rich, high-dimensional space to learn complex decision boundaries from. This likely contributed to classifiers like SVM achieving strong performance.
- Automation: No manual feature engineering is required, saving time and removing potential bias compared to handcrafted features. End-to-end training also tailor features directly to the classification task (Huang, (2016).).

5.4.2 CLAHE-4

While a comprehensive review of the literature was conducted to identify studies that employed CLAHE in stroke prediction as a pre-processing technique, no relevant studies were found. However, too many researchers studied the priority of image enhancement before AI modeling and reported that the selection of optimum pre-processing technique that appropriate to specific type of dataset will maximize the analysis and training performance on medical images this is due to:

- The enhancement of the image contrast, CLAHE works by improving the CT images contrast on a local scale that assessed quantitatively by EME. EME serves as an indicator of contrast variation within an image by comparing the local contrast across the image to a reference level. An increase in EME values after applying CLAHE indicates a successful enhancement of contrast at specific scales within the image. This enhancement can be critical for the subsequent application of AI modeling, as higher contrast can lead to more accurate feature extraction and therefore better model performance. Our results demonstrate a significant improvement in contrast, as evidenced by higher EME measurements post-CLAHE application. These findings suggest that the use of CLAHE for pre-processing CT images can lead to a more pronounced and detailed representation of the anatomical structures, thus enhancing the quality of the input data for AI modeling in thesis work.
- Visual representation standardization for automated pattern recognition. In this study dataset where B-NECT used to train the model, stroke detection especially in hyper-acute and acute stages, the low contrast and the large dynamic range that limited the visual representation of stroke lesion is improved by optimizing the CLAHE at Clip Limit 4, which led to improved model performance as noticed from the results (Shanmugavadivu & Balasubramanian, 2014).
- The reduction of noise and inconsistences in raw images that lead to mislead model during training. PSNR is a widely-used metric for measuring the quality of reconstructed images by comparing the similarity between the original image and the pre-processed image. It specifically quantifies the ratio between the maximum possible signal of an image and the power of corrupting noise that affects its fidelity. A higher PSNR value generally indicates a lower level of noise and, thus, a higher quality image.

This reduction of noise and improvement of contrast have direct implications on the modeling performance for several reasons. First, noise reduction aids in mitigating the risk of overfitting, where a model learns noise as if it were a part of the signal. This is essential for enhancing the generalizability of the AI model. Second, improving image contrast via pre-processing like CLAHE can lead to more accurate segmentation and feature extraction, critical steps in the workflow of image analysis using DenseNet-121. As a result, models built on pre-processed images are typically more accurate and reliable in their predictions, which is highly desirable in clinical diagnostics and related research fields.

Therefore, the combination of PSNR and EME provides a comprehensive assessment of the quality and applicability of pre-processed images for use in AI models. By including these metrics in our analysis, we can confirm that CLAHE significantly enhances the utility of CT images for the development of robust AI systems in medical imaging.

Chapter six: Conclusion and Recommendation

6.1 Conclusion

In this study, the examination of the DenseNet-121 deep learning model's capacity for feature extraction in stroke identification is conducted in two distinct stages. Initially, the original dataset, without the application of the CLAHE image enhancement technique, is utilized to ascertain the impact of DenseNet-121 on the performance of various classifiers. The results from this first stage indicated that the DenseNet-121 significantly improved the model's performance across all considered machine learning classifiers—RF, SVM, and LR. It resulted in high metrics across the board; RF achieved an accuracy between 0.90-0.94, SVM between 0.94-0.97, and LR was found to be between 0.90-0.97, with all classifiers exhibiting high precision, recall, f1-score, and AUC.

Upon entering the second stage of the study, CLAHE applied to enhance the CT images, referred to as SPEM, to evaluate further gains in model functionality. The application of CLAHE led to noticeable improvements in performance metrics for all models, with accuracy rates for classifiers between 0.98-0.99, paired with enhanced precision, recall, f1-score, and AUC, signifying the effectiveness of image enhancement in boosting diagnostic precision.

Among the classifiers models performed very well, however, LR distinguished itself by demonstrating superior performance in terms of computational efficiency, with elapsed classification times ranging from 0.03 to 0.4 seconds, outpacing both SVM and RF. Of interest, the performance time for the RF and SVM classifier was variable, changing according to the stroke stage, highlighting the importance of selecting the appropriate model depending on the specific requirements of real-time applications.

In conclusion, the first stage of the study established the formidable baseline capabilities of DenseNet-121 for feature extraction across various classifiers, while the second stage with the CLAHE-enhanced dataset not only substantiated the enhancement in diagnostic accuracy but also underscored the advantage of LR in achieving rapid real-time classification.

Through this pioneering work, we not only contribute to the advancement of medical imaging analytics but also bring hope for more effective stroke management protocols, underscoring the pivotal role of integrated technologies in the future of healthcare.

6.2 Recommendation

Clinical Validation is recommended to increase the efficacy and generalizability of the developed model in real-word clinical application. However, this process required availability of sufficiently large and diverse dataset especially for hyper-acute stage in which dataset is limited which can be achieved by:

I. Extend the period of data collection.
II. Involving a variety of hospitals to include governmental, non-governmental (NGOs) and private hospitals in Palestine. By this expanded involvement, strong generalization achieved. Moreover, this diversity is crucial to develop the model to deal with different patient cases, stroke severity and imaging protocols.

From findings achieved in this model it is recommended to:

I. Study the ability to employ the model to include more brain diseases as brain tumours, haemorrhages, etc.
II. Study the ability of model employment in other diseases as breast cancer, lung cancer, liver cancer classifications in different medical imaging modalities as MRI, ultrasound, mammography, etc.
III. Study the impact of CLAHE as image processing technique in the sensitivity of diseases detection in different medical imaging techniques.
IV. Extend the study to include different dataset types including patient age, gender and history, etc. that enables to develop the model for diagnosis and treatment planning.

6.3 Study Limitation

While the results of this study provide a valuable high-performance model, several limitations impeded the research progress:

- In the collection stage, there were some difficulties reaching the data site due to the unfortunate frequent occupational incursions in West Bank especially in Nablus and Jenin. So that, this cost a large effort and a long period for data collection.
- Limited data size especially in the hyper-acute and sub-acute stages, larger and more diverse datasets are needed to validate our results and provide more conclusive evidence.

- As a retrospective study, there were missing data including DW-MRI, B-NECT images and reports.
- One pertinent limitation of the current study is its patient representation, which is predominantly sourced from three private hospitals. This selective scope restricts our findings as it doesn't encompass the potentially diverse socioeconomic and health profiles of patients admitted to governmental hospitals. Including patients from public health facilities could have offered a more comprehensive understanding and increased the generalizability of our results. Future research should aim to incorporate a broader patient demographic, potentially including those from governmental hospitals, to better reflect the entire population and enhance the study's applicability to a wider clinical context.

6.4 Future Studies

The proposes of this work showed a novel and powerful promising advancements and innovations in healthcare toward the prediction, classification, and management of ischemic strokes especially in early ischemic stages in which the sensitivity of stroke detection visually is limited. By utilizing the robustness and strength of DenseNet-121 in extraction of radiomic features from pre-processed B-NECT by CLAHE, coupled with advanced classifiers (SVM, RF, and LR) to predict and classify ischemic stroke stages. As we look towards what the future holds, subsequent investigations were mapped out to expand the power and scope of this diagnostic instrument.

Through this process, the mission is not only to validate the model but also to alter it into globally online available tool that provide the healthcare professionals the ability to save the important time for save lives accurately:

1. Model deployment: By transition the model from the research phase to user-centric web platform in which health professionals can access in real-time.
2. Interface and Accessibility: By Creating a secure and intuitive interface that allows the easy upload and analysis of CT images, as well as a good integration with the healthcare IT systems.
3. Algorithm enhancement for stroke segmentation and localization and lesion size assessment by enhancing the algorithm to not only detect and predict ischemic stroke

but also to provide precise segmentation of the lesion as represented in the Figure 35, which is critical for determining the stroke's extent and affected brain regions.

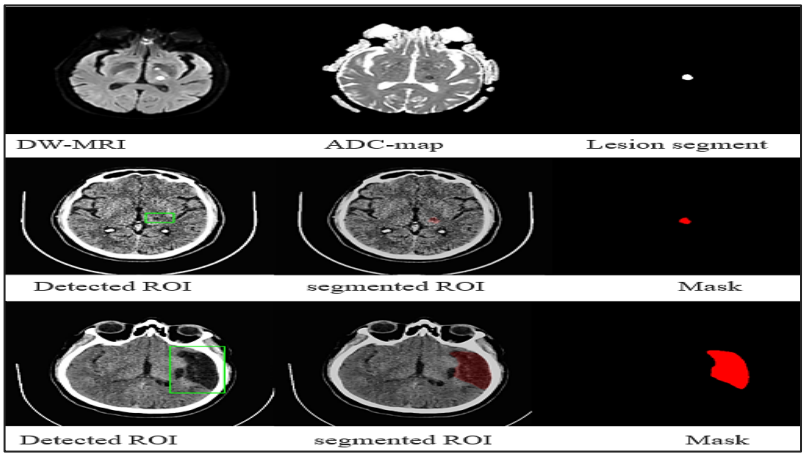


Figure 35: Hybrid Segmentation Model Deployment.

4. Expanding dataset variety by including more dataset from different CT scanners types from different hospitals and expanded patient demographics for further studies to increase the robustness and generalizability of the model.
5. Real-Time Feedback for Emergency Settings that offers immediate insights to the medical staff in the emergency states which improve the speed and accuracy of stroke interventions.
6. Clinical Trials and Validation: By Conduct extensive clinical trials involved complementary team of highly experienced medical specialists including radiologists and neurologists with data scientists and engineers, software developers, and IT specialists.
7. Developing an Open-Source Framework which enable the researchers and developers to improve the application over the time.
8. Ethical and Privacy Considerations: by Addressing any ethical concerns and ensure privacy by employing strict.

Ethical Considerations

I. To perform the study, the proposal was presented to Al-Quds University - Faculty of Health professions review board to gain agreement and permission.

II. The three mentioned private hospitals gave the permission to conduct the study in their radiology departments.

III. The privacy of the collected data was maintained. By not mentioning names, patients ID number, or codes that could be referred to obtain detailed about personal information to a definite patient. - The study should not present any conflict of interest.

References

Abedi, V., Goyal, N., Tsivgoulis, G., Hosseinichimeh, N., Hontecillas, R., Bassaganya-Riera, J., Elijovich, L., Metter, J. E., Alexandrov, A. W., Liebeskind, D. S., Alexandrov, A. V., & Zand, R. (2017). Novel Screening Tool for Stroke Using Artificial Neural Network. *Stroke*, *48*(6), 1678-1681. https://doi.org/10.1161/strokeaha.117.017033

Abedi, V., Khan, A., Chaudhary, D., Misra, D., Avula, V., Mathrawala, D., Kraus, C., Marshall, K. A., Chaudhary, N., & Li, X. (2020). Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Therapeutic advances in neurological disorders*, *13*, 1756286420938962.

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput Biol Med*, *100*, 270-278. https://doi.org/10.1016/j.compbiomed.2017.09.017

Aizerman, A. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, *25*, 821-837.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on Medical Imaging*, *35*(5), 1207-1216.

Barber PA, D. A., Zhang J, Buchan AM. (2000). Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet ASGJT*, *355(9216)*, 1670–1674.

Barber PA, D. A., Zhang J, Buchan AM. (2000). Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. *Alberta Stroke Programme Early CT Score. Lancet.*, *355*, 1670–1674.

Barber, P. A., Demchuk, A. M., Zhang, J., & Buchan, A. M. (2000). Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *The Lancet*, *355*(9216), 1670-1674.

Barber PA, H. M., Eliasziw M, Demchuk AM, Pexman JHW, Hudon ME, et al. (2005). Imaging of the brain in acute ischaemic stroke: comparison of computed tomography

and magnetic resonance diffusion-weighted imaging. *Neurol Neurosurg Psychiatry*, *76*, 76:1528. https://doi.org/10.1136/jnnp.2004.059261

Becks, M. J., Manniesing, R., Vister, J., Pegge, S. A. H., Steens, S. C. A., van Dijk, E. J., Prokop, M., & Meijer, F. J. A. (2019). Brain CT perfusion improves intracranial vessel occlusion detection on CT angiography. *J Neuroradiol*, *46*(2), 124-129. https://doi.org/10.1016/j.neurad.2018.03.003

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24-31.

Bengio, Y., Delalleau, O., & Roux, N. (2005). The curse of highly variable functions for local kernel machines. *Advances in neural information processing systems*, *18*.

Berkhemer, O. A., Fransen, P. S., Beumer, D., Van Den Berg, L. A., Lingsma, H. F., Yoo, A. J., Schonewille, W. J., Vos, J. A., Nederkoorn, P. J., & Wermer, M. J. (2015). A randomized trial of intraarterial treatment for acute ischemic stroke. *New England Journal of Medicine*, *372*(1), 11-20.

Bishop, C. (2006). Pattern recognition and machine learning. *Springer google schola*, *2*, 531-537.

Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. *Advances in neural information processing systems*, *20*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Byrne, D., Sugrue, G., Stanley, E., Walsh, J. P., Murphy, S., Kavanagh, E. C., & MacMahon, P. J. (2017). Improved Detection of Anterior Circulation Occlusions: The "Delayed Vessel Sign" on Multiphase CT Angiography. *AJNR Am J Neuroradiol*, *38*(10), 1911-1916. https://doi.org/10.3174/ajnr.A5317

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS computational biology*, *10*(12), e1003963.

Campbell, B. C., Christensen, S., Levi, C. R., Desmond, P. M., Donnan, G. A., Davis, S. M., & Parsons, M. W. (2012). Comparison of computed tomography perfusion and magnetic resonance imaging perfusion-diffusion mismatch in ischemic stroke. *Stroke*, *43*(10), 2648-2653. https://doi.org/10.1161/strokeaha.112.660548

Chen, H., Qi, X., Yu, L., & Heng, P.-A. (2016). DCAN: deep contour-aware networks for accurate gland segmentation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,

Cheon, S., et al. (2019). The Use of Deep Learning to Predict Stroke Patient Mortality. *International Journal of Environmental Research and Public Health*, *11*.

Chiang, P.-L., Lin, S.-Y., Chen, M.-H., Chen, Y.-S., Wang, C.-K., Wu, M.-C., Huang, Y.-T., Lee, M.-Y., Chen, Y.-S., & Lin, W.-C. (2022). Deep Learning-Based Automatic Detection of ASPECTS in Acute Ischemic Stroke: Improving Stroke Assessment on CT Scans. *Journal of Clinical Medicine*, *11*(17), 5159.

Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2019). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput Biol Med*, *115*, 103487. https://doi.org/10.1016/j.compbiomed.2019.103487

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, *31*(4-5), 198-211.

Donahue, J., & Wintermark, M. (2015). Perfusion CT and acute stroke imaging: foundations, applications, and literature review. *J Neuroradiol*, *42*(1), 21-29. https://doi.org/10.1016/j.neurad.2014.11.003

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis* (Vol. 3). Wiley New York.

Dzialowski I, K. E., Goericke S, Doerfler A, Forsting M, von Kummer R. (2007). Ischemic brain tissue water content: CT monitoring during middle cerebral artery occlusion and reperfusion in rats. *Radiology*, *243*, 720–726.

Emberson J, L. K., Lyden P, Blackwell L, Albers G, Bluhmki E, et al. (2014). Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet*, *384*, 1929–1935.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1915-1929.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Scene parsing with multiscale feature learning, purity trees, and optimal covers. *arXiv preprint arXiv:1202.2160*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Fawzi, A., Achuthan, A., & Belaton, B. (2021). Adaptive Clip Limit Tile Size Histogram Equalization for Non-Homogenized Intensity Images. *IEEE Access*, *9*, 164466-164492.

Finck, T., Schinz, D., Grundl, L., Eisawy, R., Yiğitsoy, M., Moosbauer, J., Zimmer, C., Pfister, F., & Wiestler, B. (2022). Automated Detection of Ischemic Stroke and Subsequent Patient Triage in Routinely Acquired Head CT. *Clinical Neuroradiology*, *32*(2), 419-426.

García-Díaz, P., Sanchez-Berriel, I., Martínez-Rojas, J. A., & Diez-Pascual, A. M. (2020). Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. *Genomics*, *112*(2), 1916-1925.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, *31*(14), 2225-2236.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gu J, W. Z., Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354.

Hadsell, R., Sermanet, P., Ben, J., Erkan, A., Scoffier, M., Kavukcuoglu, K., Muller, U., & LeCun, Y. (2009). Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, *26*(2), 120-144.

Hanning, H. C. K. P. B. S. G. B. A. K. J. N. T. R. J. F. U. (2020). Posterior circulation stroke: machine learning-based detection of early ischemic changes in acute non-contrast CT scans. *Journal of Neurology*, *267*, 2632–2641.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediciton.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, *21*(9), 1263-1284.

Hebli, A., & Gupta, S. (2017). Brain tumor prediction and classification using support vector machine. 2017 international conference on advances in computing, communication and control (ICAC3),

Heo, J. N., et al. (2019). Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke*, *vol. 50, no. 5*, 1263–1265.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Huang, G., Liu, Z., & Weinberger, K. Q. . ((2016).). Densely Connected Convolutional Networks. *ArXiv. /abs/1608.06993*.

Isa, I. S., Sulaiman, S. N., Mustapha, M., & Karim, N. K. A. (2017). Automatic contrast enhancement of brain MR images using Average Intensity Replacement based on Adaptive Histogram Equalization (AIR-AHE). *Biocybernetics and Biomedical Engineering*, *37*(1), 24-34.

JE., T. (1996). The evolution of surgery for the treatment and prevention of stroke. The Willis Lecture., *27(8)*(Stroke), 1427– 1434. https://doi.org/10.1161/01.STR.27.8.1427

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition workshops,

Johnson, W., Onuma, O., Owolabi, M., and Sachdev, S. (2016). Stroke: a global response is needed. *Bulletin of the World Health Organization*, 94:634–634A.

Josephson SA, H. N., Johnston SC. (2006). NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis*, *22(5-6)*, 389–395.

Jovin TG, C. A., Cobo E, de Miquel MA, Molina CA, Rovira A, et al. (2015). Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*, *372*, 2296–2306.

Jun Lu, Y. Z., Wenzhi Lv, Hongquan Zhu, Tian Tian, Su Yan, Yan Xie, Di Wu, Yuanhao Li, Yufei Liu, Luyue Gao, Wei Fan, Yan Nan, Shun Zhang, Xiaolong Peng, Guiling Zhang, Wenzhen Zhu. (2022). Identification of early invisible acute ischemic stroke in non-contrast computed tomography using two-stagedeep-learning model. *Theranostics*, *12*(12), 5564- 5573.

Kansal, S., Purwar, S., & Tripathi, R. K. (2018). Image contrast enhancement using unsharp masking and histogram equalization. *Multimedia Tools and Applications*, *77*, 26919-26938.

Kasner, S. E. (2006). Clinical interpretation and use of stroke scales. *Lancet Neurol*, *5*(7), 603-612. https://doi.org/10.1016/s1474-4422(06)70495-1

Khatib, R., Jawaadah, A. M., Khammash, U., Babiker, A., Huffman, M. D., & Prabhakaran, S. (2018). Presentation, Management, and Outcomes of Acute Stroke in Palestine. *J Am Heart Assoc*, *7*(22), e010778. https://doi.org/10.1161/jaha.118.010778

Kidwell, C. S., Alger, J. R., & Saver, J. L. (2003). Beyond mismatch: evolving paradigms in imaging the ischemic penumbra with multimodal magnetic resonance imaging. *Stroke*, *34*(11), 2729-2735.

Kim, H., Golub, G. H., & Park, H. (2005). Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics*, *21*(2), 187-198.

Kniep, H. C., Sporns, P. B., Broocks, G., Kemmling, A., Nawabi, J., Rusche, T., Fiehler, J., & Hanning, U. (2020). Posterior circulation stroke: machine learning-based detection of early ischemic changes in acute non-contrast CT scans. *Journal of Neurology*, *267*, 2632-2641.

Konstas AA, e. a. (2009). Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke part 1: theoretic basis. *AJNR*, *30(4)*(Am J Neuroradiol), 662—668.

Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., & Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*, *35*, 303-312. https://doi.org/10.1016/j.media.2016.07.007

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial Intelligence in Precision Cardiovascular Medicine. *J Am Coll Cardiol*, *69*(21), 2657-2664. https://doi.org/10.1016/j.jacc.2017.03.571

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Krutsch, R., & Tenorio, D. (2011). Histogram equalization. *Freescale Semiconductor, Document Number AN4318, Application Note*.

Lebedev, G., Meshcheryakova, A., Kurenkov, V., Kluganov, V., Sologubov, A., Logacheva, N., Radzievskiy, G., & Klimenko, H. (2020). Application of artificial intelligence methods to recognize pathologies on photographs of fundus. *Procedia Computer Science*, *176*, 1823-1828.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., & Hubbard, W. (1990). Proc. Advances in Neural Information Processing Systems. In.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

LeCun Y, B. Y., Hinton. (2015). G. Deep learning. *Nature*, *521(7553)*, 436–444.

Lee, E. J., et al. (2017). Deep into the Brain: Artificial Intelligence in Stroke Imaging. *Journal of Stroke*, *vol.19 no. 3*, 277–285.

Lee, H., Lee, E.-J., Ham, S., Lee, H.-B., Lee, J. S., Kwon, S. U., Kim, J. S., Kim, N., & Kang, D.-W. (2020). Machine learning approach to identify stroke within 4.5 hours. *Stroke*, *51*(3), 860-866.

Leung, M. K., Xiong, H. Y., Lee, L. J., & Frey, B. J. (2014). Deep learning of the tissue-regulated splicing code. *Bioinformatics*, *30*(12), i121-i129.

Lev, M. H., Farkas, J., Gemmete, J. J., Hossain, S. T., Hunter, G. J., Koroshetz, W. J., & Gonzalez, R. G. (1999). Acute stroke: improved nonenhanced CT detection—benefits of soft-copy interpretation by using variable window width and center level settings. *Radiology*, *213*(1), 150-155.

Li, H., Chen, J., Lu, H., & Chi, Z. (2017). CNN for saliency detection with low-level feature integration. *Neurocomputing*, *226*, 212-220.

Li, J., Liang, P., Shen, L., Dong, B., ... & Zhang, C. . (2018). A densely connected CNN with depth-wise dilated convolution for automatic pulmonary nodule detection. *Neurocomputing*(275), 12-22.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., & Heng, P.-A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE transactions on Medical Imaging*, *37*(12), 2663-2674.

Liao, C.-C., Xiao, F., Wong, J.-M., & Chiang, I.-J. (2006). A simple genetic algorithm for tracing the deformed midline on a single slice of brain CT using quadratic Bézier

curves. Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06),

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Lightfoot, C. B., Abraham, R. J., Mammen, T., Abdolell, M., Kapur, S., & Abraham, R. J. (2009). Survey of radiologists' knowledge regarding the management of severe contrast material–induced allergic reactions. *Radiology*, *251*(3), 691-696.

Lin CB, P. E., Smith EE, et al. (2012). Emergency medical service hospital prenotification is associated with improved evaluation and treatment of acute ischemic stroke. *Circ Cardiovasc Qual Outcomes*, *5(04)*, 514–522.

Lin, S.-Y., Chiang, P.-L., Chen, P.-W., Cheng, L.-H., Chen, M.-H., Chang, P.-C., Lin, W.-C., & Chen, Y.-S. (2022). Toward automated segmentation for acute ischemic stroke using non-contrast computed tomography. *International journal of computer assisted radiology and surgery*, *17*(4), 661-671.

Lisowska, A., O'Neil, A., Dilys, V., Daykin, M., Beveridge, E., Muir, K., Mclaughlin, S., & Poole, I. (2017). Context-aware convolutional neural networks for stroke sign detection in non-contrast CT scans. Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21,

Lo, C.-M., Hung, P.-H., & Hsieh, K. L.-C. (2019). Computer-aided detection of hyperacute stroke based on relative radiomic patterns in computed tomography. *Applied Sciences*, *9*(8), 1668.

Lo, S.-C. B., Chan, H.-P., Lin, J.-S., Li, H., Freedman, M. T., & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural networks*, *8*(7-8), 1201-1214.

Lo, S.-C. B., Lin, J.-S., Freedman, M. T., & Mun, S. K. (1993). Computer-assisted diagnosis of lung nodule detection using artificial convoultion neural network. Medical Imaging 1993: Image Processing,

Lombardi, S., Riva, L., Patassini, M., Remida, P., Capraro, C., Canonico, F., Franzesi, C. T., & Ippolito, D. (2018). "Hyperdense artery sign" in early ischemic stroke: diagnostic value of model-based reconstruction approach in comparison with standard hybrid iterative reconstruction algorithm. *Neuroradiology*, *60*, 1273-1280.

Lyden P, R. R., Liu L, Emr M, Warren M, Marler J. (2009). National Institutes of Health Stroke Scale certification is reliable across multiple venues. *Stroke*, *40(07)*, 2507–2511.

Mainali, S., Wahba, M., & Elijovich, L. (2014). Detection of Early Ischemic Changes in Noncontrast CT Head Improved with "Stroke Windows". *ISRN Neurosci*, *2014*, 654980. https://doi.org/10.1155/2014/654980

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Maroulis, D. E., Savelonas, M. A., Karkanis, S. A., Iakovidis, D. K., & Dimitropoulos, N. (2005). Computer-aided thyroid nodule detection in ultrasound images. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05),

Mechaliski RS, B. I., Kubat M. (1998). Machine learning, data mining and knowledge discovery: methods and applications. *New York: Wiley*.

Melingi, S. B., Mojjada, R. K., Tamizhselvan, C., Surender, R., & Yazhinian, S. (2022). A self-adaptive monarch butterfly optimization (MBO) algorithm based improved deep forest neural network model for detecting and classifying brain stroke lesions. *Research on Biomedical Engineering*, *38*(2), 647-660. https://doi.org/10.1007/s42600-022-00214-2

Menard, S. (2002). *Applied logistic regression analysis*. Sage.

Menon BK, A. M., Pereira VM, et al. (2014). Optimal workflow and process-based performance measures for endovascular therapy in acute ischemic stroke: analysis of the Solitaire FR thrombectomy for acute revascularization study. . *Stroke*, *45(7)*, 2024–2029.

Menon, B. K., d'Esterre, C. D., Qazi, E. M., Almekhlafi, M., Hahn, L., Demchuk, A. M., & Goyal, M. (2015). Multiphase CT Angiography: A New Tool for the Imaging Triage of Patients with Acute Ischemic Stroke. *Radiology*, *275*(2), 510-520. https://doi.org/10.1148/radiol.15142256

Menon BK, D. A. (2009). Is acute stroke treatment heading towards a more endovascular approach? *Lancet Neurol*(8(9)), 778–779.

Menon BK, G. M. (2011). Endovascular therapy in acute ischemic stroke: where we are, the challenges we face and what the future holds. *Expert Rev Cardiovasc Ther*, *9(4)*, 473–484.

Menon BK, H. M., Eesa M, et al. (2011). Initial experience with the Penumbra Stroke System for recanalization of large vessel occlusions in acute ischemic stroke. *Neuroradiology*, *53(4)*, 261–266.

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J. (2011). Strategies for training large scale neural network language models. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding,

MOH, P. M. o. H. (2023). *Health Annual Report, Palestine 2022,* Major 10th Causes of Death by district, Palestine 2022). www.moh.ps

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Mozaffarian D, B. E., Go AS, et al. (2016). Heart Disease and Stroke Statistics-2016 Update: a report from the American Heart Association. Circulation. *Writing Group Members; American Heart Association Statistics Committee*, *133(04)*, e38–e360.

Musuka, T. D., Wilton, S. B., Traboulsi, M., & Hill, M. D. (2015). Diagnosis and management of acute ischemic stroke: speed is critical. *Cmaj*, *187*(12), 887-893. https://doi.org/10.1503/cmaj.140355

N. Tomura, K. U., A. Inugami, H. Fujita, S. Higano, and F. Shishido. (1988). Early CT finding in cerebral infarction: obscuration of the lentiform nucleus. *Radiology*, *168, no. 2*, 463–467.

Nael, K., Khan, R., Choudhary, G., Meshksar, A., Villablanca, P., Tay, J., Drake, K., Coull, B. M., & Kidwell, C. S. (2014). Six-minute magnetic resonance imaging protocol for evaluation of acute ischemic stroke: pushing the boundaries. *Stroke*, *45*(7), 1985-1991.

Ospel, J. M., Volny, O., Qiu, W., Najm, M., Kashani, N., Goyal, M., & Menon, B. K. (2020). Displaying Multiphase CT Angiography Using a Time-Variant Color Map: Practical Considerations and Potential Applications in Patients with Acute Stroke. *AJNR Am J Neuroradiol*, *41*(2), 200-205. https://doi.org/10.3174/ajnr.A6376

Pan, C.-L. C. B.-J. L. G.-R. W. T.-C. W. C.-S. Y. R.-C. S. Y.-J. (2017). An Automated Early Ischemic Stroke Detection System using CNN Deep Learning Algorithm. IEEE 8th International Conference on Awareness Science and Technology (iCAST 2017), Taichung, Taiwan.

Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE transactions on neural networks*, *22*(2), 199-210.

Parsons, M., Spratt, N., Bivard, A., Campbell, B., Chung, K., Miteff, F., O'Brien, B., Bladin, C., McElduff, P., & Allen, C. (2012). A randomized trial of tenecteplase versus alteplase for acute ischemic stroke. *New England Journal of Medicine*, *366*(12), 1099-1107.

Pedro P. Rebouc¸as Filho, R. M. S., Gabriel Bandeira Holanda, Daniel de Alencar Lima. (2017). New approach to detect and classify stroke in skull CT images via Analysis of Brain Tissue Densities. *Computer Methods and Programs in Biomedicine*.

Peixoto, S. A., & Rebouças Filho, P. P. (2018). Neurologist-level classification of stroke using a structural co-occurrence matrix based on the frequency domain. *Computers & Electrical Engineering*, *71*, 398-407.

Peng, S. J., Chen, Y. W., Yang, J. Y., Wang, K. W., & Tsai, J. Z. (2022). Automated Cerebral Infarct Detection on Computed Tomography Images Based on Deep Learning. *Biomedicines*, *10*(1). https://doi.org/10.3390/biomedicines10010122

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on Medical Imaging*, *35*(5), 1240-1251.

Peter, R., Korfiatis, P., Blezek, D., Oscar Beitia, A., Stepan-Buksakowska, I., Horinek, D., Flemming, K. D., & Erickson, B. J. (2017). A quantitative symmetry-based analysis of hyperacute ischemic stroke lesions in noncontrast computed tomography. *Medical physics*, *44*(1), 192-199.

PMOH, P. M. O. H. (2018). *REFERRAL PROTOCOL 7: MEDICAL PROCEDURES AND OTHER TESTS*. https://site.moh.ps/index/Books/BookType/2/Language/ar

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Prabhakaran, S., Ruff, I., & Bernstein, R. A. (2015). Acute stroke intervention: a systematic review. *JAMA*, *313*(14), 1451-1462.

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, *9*(3), e1301.

Przelaskowski, A., Sklinda, K., Bargieł, P., Walecki, J., Biesiadko-Matuszewska, M., & Kazubek, M. (2007). Improved early stroke detection: wavelet-based perception enhancement of computerized tomography exams. *Computers in Biology and Medicine*, *37*(4), 524-533.

Puetz V, S. P., Coutts SB, et al. . (2008). Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion. *Stroke*, *39(09)*, 2485–2490.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*, 81-106.

Rani, S., & Kumar, M. (2013). A Review of Image Contrast Enhancement Techniques.

Reboucas Filho, P. P., Sarmento, R. M., Holanda, G. B., & de Alencar Lima, D. (2017). New approach to detect and classify stroke in skull CT images via analysis of brain tissue densities. *Computer Methods and Programs in Biomedicine*, *148*, 27-43.

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, *38*, 35-44.

Rha JH, S. J. (2007). The impact of recanalization on ischemic stroke outcome: a metaanalysis. . *Stroke* 38(3), 967–973.

Ritzl, A., Meisel, S., Wittsack, H. J., Fink, G. R., Siebler, M., Mödder, U., & Seitz, R. J. (2004). Development of brain infarct volume as assessed by magnetic resonance imaging (MRI): Follow-up of diffusion-weighted MRI lesions. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *20*(2), 201-207.

Roman Peter. Panagiotis Korfiatis, D. B., and A. Oscar Beitia Irena Stepan-Buksakowska and Daniel Horinek. Kelly D. Flemming. Bradley J. Erickson. (2017). A quantitative symmetry-based analysis of hyperacute ischemic stroke lesions in noncontrast computed tomography. *American Association of Physicists in Medicine*, *2016*.

Rudkin, S., Cerejo, R., Tayal, A., & Goldberg, M. F. (2018). Imaging of acute ischemic stroke. *Emerg Radiol*, *25*(6), 659-672. https://doi.org/10.1007/s10140-018-1623-x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536.

Russakovsky O, D. J., Su H et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis 115*, 211–252.

Sahiner, B., Chan, H.-P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D., & Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE transactions on Medical Imaging*, *15*(5), 598-610.

Sailasya, G., & Kumari, G. L. A. (2021). Analyzing the performance of stroke prediction using ML classification algorithms. *International Journal of Advanced Computer Science and Applications*, *12*(6).

Salvi, M., Acharya, U. R., Molinari, F., & Meiburger, K. M. (2021). The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, *128*, 104129.

Saver, J. L. (2017). Penumbral salvage and thrombolysis outcome: a drop of brain, a week of life. *Brain*, *140*(3), 519-522. https://doi.org/10.1093/brain/awx020

Saver JL, G. M., van der Lugt A, et al; HERMES Collaborators. (2016). Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA*, *316(12)*, 1279–1288.

Schaefer, P. W., Ozsunar, Y., He, J., Hamberg, L. M., Hunter, G. J., Sorensen, A. G., Koroshetz, W. J., & Gonzalez, R. G. (2003). Assessing tissue viability with MR diffusion and perfusion imaging. *American Journal of Neuroradiology*, *24*(3), 436-443.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Schölkopf, B., Smola, A. J., & Bach, F. (2002). Learning with kernels: support vector machines, Regularization. *Optimization, and Beyond. MIT press*, *1*(2).

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

Shanmugavadivu, P., & Balasubramanian, K. (2014). Thresholded and Optimized Histogram Equalization for contrast enhancement of images. *Computers & Electrical Engineering*, *40*(3), 757-768. https://doi.org/https://doi.org/10.1016/j.compeleceng.2013.06.013

Sharath KG, N. C. (2018). Acute ischemic stroke: a review of imaging, patient selection, and management in the endovascular era. Part II: Patient selection, endovascular thrombectomy, and postprocedure management. *J Clin Interventional Radiology Isvir*, *02(03)*, 1975–1980.

Shraddha Mainali, M. W., and Lucas Elijovich. (2014). Detection of Early Ischemic Changes in Noncontrast CT Head Improved with ‘‘Stroke Windows’’. *ISRN Neuroscience*, *2014*, 4 pages.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Smajlović, D., & Sinanović, O. (2004). Sensitivity of the neuroimaging techniques in ischemic stroke. *Med Arh*, *58*(5), 282-284.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427-437.

Srinivasan, A., Goyal, M., Azri, F. A., & Lum, C. (2006). State-of-the-Art Imaging of Acute Stroke. *RadioGraphics*, *26*(suppl_1), S75-S95. https://doi.org/10.1148/rg.26si065501

Statnikov, A., Wang, L., & Aliferis, C. (2008). Statnikov A, Wang L, Aliferis CFA comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 9: 319. *BMC bioinformatics*, *9*, 319. https://doi.org/10.1186/1471-2105-9-319

Syu-Jyun Peng, Y.-W. C., Jing-Yu Yang, Kuo-Wei Wang and Jang-Zern Tsai. (2022). Automated Cerebral Infarct Detection on Computed Tomography Images Based on Deep Learning. *Biomedicines*, *10*, 122.

Tan, T., Sim, k. s., Tso, C. P., & Chong, A. (2012). Contrast enhancement of computed tomography images by adaptive histogram equalization-application for improved ischemic stroke detection. *International Journal of Imaging Systems and Technology*, *22*, 153-160. https://doi.org/10.1002/ima.22016

Tang, F. H., Ng, D. K., & Chow, D. H. (2011). An image feature approach for computer-aided detection of ischemic stroke. *Comput Biol Med*, *41*(7), 529-536. https://doi.org/10.1016/j.compbiomed.2011.05.001

Tei H, U. S., Usui T, Ohara K. (2010). Posterior circulation ASPECTS on diffusion-weighted MRI can be a powerful marker for predicting functional outcome. *J Neurol*, *257(05)*, 767–773.

Thurnher, M. (2008). *Imaging in Acute Stroke*.

Tyan, Y.-S., Wu, M.-C., Chin, C.-L., Kuo, Y.-L., Lee, M.-S., & Chang, H.-Y. (2014). Ischemic Stroke Detection System with a Computer-Aided Diagnostic Ability Using an Unsupervised Feature Perception Enhancement Method. *International Journal of Biomedical Imaging*, *2014*, 947539. https://doi.org/10.1155/2014/947539

Vagal, A., Foster, L. D., Menon, B., Livorine, A., Shi, J., Qazi, E., Yeatts, S. D., Demchuk, A. M., Hill, M. D., Tomsick, T. A., & Goyal, M. (2016). Multimodal CT Imaging: Time to Treatment and Outcomes in the IMS III Trial. *AJNR Am J Neuroradiol*, *37*(8), 1393-1398. https://doi.org/10.3174/ajnr.A4751

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, *10*(5), 988-999.

Vogelsang, F., Weiler, F., Dahmen, J., Kilbinger, M. W., Wein, B. B., & Guenther, R. W. (1998). Detection and compensation of rib structures in chest radiographs for diagnostic assistance. Medical Imaging 1998: Image Processing,

Volny, O., Cimflova, P., Kadlecova, P., Vanek, P., Vanicek, J., Menon, B. K., & Mikulik, R. (2017). Single-Phase Versus Multiphase CT Angiography in Middle Cerebral Artery Clot Detection-Benefits for Less Experienced Radiologists and Neurologists. *J Stroke Cerebrovasc Dis*, *26*(1), 19-24. https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.08.023

Waibel, A., Sawai, H., & Shikano, K. (1989). Modularity and scaling in large phonemic neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *37*(12), 1888-1898.

Wang, L., Qiao, Y., & Tang, X. (2013). Motionlets: Mid-level 3d parts for human motion recognition. Proceedings of the ieee conference on computer vision and pattern recognition,

Wang, X., Jin, Y., Long, M., Wang, J., & Jordan, M. (2020). *Transferable Normalization: Towards Improving Transferability of Deep Neural Networks*.

Wang, Z., Xie, J., Tang, T.-Y., Zeng, C.-H., Zhang, Y., Zhao, Z., Zhao, D.-L., Geng, L.-Y., Deng, G., Zhang, Z.-J., Ju, S.-H., & Teng, G.-J. (2020). Collateral Status at Single-Phase and Multiphase CT Angiography versus CT Perfusion for Outcome Prediction in Anterior Circulation Acute Ischemic Stroke. *Radiology*, *296*(2), 393-400. https://doi.org/10.1148/radiol.2020192029

Warach, S., Chien, D., Li, W., Ronthal, M., & Edelman, R. (1992). Fast magnetic resonance diffusion-weighted imaging of acute human stroke. *Neurology*, *42*(9), 1717-1717.

Wardlaw, J. M., Farrall, A. J., Perry, D., von Kummer, R., Mielke, O., Moulin, T., Ciccone, A., & Hill, M. (2007). Factors influencing the detection of early CT signs of cerebral ischemia: an internet-based, international multiobserver study. *Stroke*, *38*(4), 1250-1256. https://doi.org/10.1161/01.Str.0000259715.53166.25

Wei, D., Sahiner, B., Chan, H.-P., & Petrick, N. (1995). Detection of masses on mammograms using a convolution neural network. 1995 international conference on acoustics, speech, and signal processing,

Wu Qiu, P. H. K., PhD. Ericka Teleg, MD. Johanna M. Ospel, MD. Sung Il Sohn, PhD, MD. Mohammed Almekhlafi, MD. Mayank Goyal, MD.Michael D. Hill, MSc, MD. Andrew M. Demchuk, MD. Bijoy K. Menon, MD, MSc. (2020). Machine Learning for Detecting Early Infarction in Acute Stroke with Non–Contrast-enhanced CT. *Radiology*, *00:1–7*.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., Hua, Y., Gueroussov, S., Najafabadi, H. S., & Hughes, T. R. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*(6218), 1254806.

Yang, C. Y., Chen, Y. F., Lee, C. W., Huang, A., Shen, Y., Wei, C., & Liu, H. M. (2008). Multiphase CT angiography versus single-phase CT angiography: comparison of image quality and radiation dose. *AJNR Am J Neuroradiol*, *29*(7), 1288-1295. https://doi.org/10.3174/ajnr.A1073

Yoo, A. J., Berkhemer, O. A., Fransen, P. S. S., van den Berg, L. A., Beumer, D., Lingsma, H. F., Schonewille, W. J., Sprengers, M. E. S., van den Berg, R., van Walderveen, M. A. A., Beenen, L. F. M., Wermer, M. J. H., Nijeholt, G., Boiten, J., Jenniskens, S. F. M., Bot, J. C. J., Boers, A. M. M., Marquering, H. A., Roos, Y., . . . Majoie, C. (2016). Effect of baseline Alberta Stroke Program Early CT Score on safety and efficacy of intra-arterial treatment: a subgroup analysis of a randomised phase 3 trial (MR CLEAN). *Lancet Neurol*, *15*(7), 685-694. https://doi.org/10.1016/s1474-4422(16)00124-1

Zhang, G. P. (2000). Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *30*(4), 451-462.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on Medical Imaging*, *20*(1), 45-57.

Zhao, Y. H., Yuan, B., Chen, J., Feng, D. H., Zhao, B., Qin, C., & Chen, Y. F. (2013). Endothelial progenitor cells: therapeutic perspective for ischemic stroke. *CNS neuroscience & therapeutics*, *19*(2), 67-75.

نموذج الذكاء الاصطناعي الهجين للتنبؤ والتصنيف عبر جميع مراحل السكتة الدماغية الإقفارية الصور المقطعية المحوسبة غير المعززة بالمادة الظليلية

إعداد الطالب: إبراهيم بسام إبراهيم قديح

إشراف: د. رضوان القصراوي

الملخص

تعتبر السكتة الدماغية مشكلة صحية عالمية كبرى، حيث تؤدي إلى إعاقة كبيرة وقد تؤدي إلى حدوث الوفيات، و بحسب تقارير منظمة الصحة العالمية (WHO) يصاب حوالي 16 مليون شخص سنويًا بهذا المرض. إن الاستجابة السريعة لهذا النوع من الأمراض أمرًا بالغ الأهمية للتخفيف من تلف الدماغ وتحسين النتائج العلاجية للمرضى. السكتات الدماغية، والتي يتم تصنيفها في المقام الأول إلى نوعين وهما الجلطات الإقفارية والجلطات النزيفية، حيث تختلف في آلية الحصول ويمكن أن تتأثر بعوامل الخطر القابلة للمعالجة. وفي مناطق مثل فلسطين، التي تعتبر من الدول ذات الموارد الاقتصادية المحدودة، تعد السكتة الدماغية سببًا شائعًا للوفاة. تتفاقم التحديات التشخيصية بسبب محدودية فحوصات الدماغ المقطعية غير المعززة بالمادة الظليلية (B-NECT)، والتي تختلف في فعاليتها بناءً على مرحلة السكتة الدماغية.

تقدم هذه الدراسة إطارًا جديدًا قائمًا على الذكاء الاصطناعي، باستخدام نموذج تحسين دقة السكتة الدماغية (SPEM) الذي تم تصميمه، و الذي يستخدم تقنيات معالجة الصور والتعلم العميق والتعلم الآلي لتعزيز تصنيف مراحل السكتة الدماغية في صور B-NECT. يدمج هذا النموذج المختلط في الوقت الفعلي معادلة الرسم البياني التكيفي المحدودة للتباين (CLAHE) للمعالجة المسبقة، مع إجراء استخراج الميزات من خلال الشبكات التلافيفية المتصلة بكثافة-121 (DenseNet-121). يتم إجراء التصنيف باستخدام آلة ناقل الدعم (SVM)، والغابات العشوائية (RF)، والانحدار اللوجستي (LR)، مع التركيز على تحديد الطريقة الأكثر فعالية بناءً على مقاييس الأداء المختلفة.

تشير النتائج إلى أداء استثنائي لـ SPEM، خاصة عند دمج DenseNet-121 مع المصنف LR. والجدير بالذكر أنه في المرحلة شديدة الحدة، حقق النموذج دقة قدرها 0.9957، و إحكام قدره 0.9914، و تقييم مميز في مقياس منطقة تحت منحنى خصائص تشغيل المستقبِل (AUC) حيث بلغت 0.9999، مع وقت معالجة وتصنيف يبلغ 0.04 ثانية فقط. وقد حافظ مصنف LR على أداء عالٍ مماثل عبر مراحل السكتة الدماغية الأخرى. تسلط هذه النتائج الضوء على إمكانات هذا النموذج المعزز بالذكاء الاصطناعي في تسهيل اتخاذ قرارات سريرية أسرع وأكثر دقة لعلاج السكتة الدماغية في المراحل المبكرة.

يُظهر النموذج الهجين نتائج واعدة في توقع وتصنيف السكتات الدماغية ويمكن أن يؤثر بشكل كبير على الممارسة السريرية عند إجراء مزيد من البحث والتحقق من صحته على مجموعات بيانات أكبر، وتعزيز إمكانية التفسير والتكامل في سير العمل السريري.

Appendix A: Letter of Research Ethical Committee (REC) approval

Al-Quds University
Jerusalem
Deanship of Scientific Research

بسم الله الرحمن الرحيم

جامعة القدس
القدس
عمادة البحث العلمي

Research Ethics Committee
Committee's Decision Letter

Date: June 4, 2023

Ref No: 300/REC/2023

Dears Dr. Radwan Qasrawi, Dr. Diala Abu Al-Halawa, Mr. Ibrahem Qdaih,

Thank you for submitting your application for research ethics approval. After reviewing your application entitled "Hybrid Machine Learning Technique for Improving Early Detection and Classification of Non-enhanced Medical Images of Brain Strokes", the Research Ethics Committee confirms that your application is in accordance with the research ethics guidelines at Al-Quds University.

We would appreciate receiving a copy of your final research report/ publication.

Thank you again and wish you a productive research that serves the best interests of your subjects.

PS: This letter will be valid for two years.

Sincerely,

Suheir Ereqat, PhD
Associate Professor of Molecular Biology

Research Ethics Committee Chair

Cc. Prof. Imad Abu Kishek - President
Cc. Members of the committee
Cc. file

Abu-Dies, Jerusalem P.O.Box 20002
Tel-Fax: #970-02-2791293
research@admin.alquds.edu
أبوديس، القدس ص.ب. 20002
تلفاكس: #970-02-2791293

Appendix B: Models Evaluation Performance of the Original Dataset Summery

Table 3: Performance Evaluation of the Original Data Using Different Machine Learning Classifiers, with respect to DenseNet121 Deep Learning Model

Brain Ischemic Stroke Prediction (without enhancement)							
Stages	**Classifier**	**Accuracy**	**Precision**	**Recall**	**F1**	**AUC**	**Time(S)**
Hyperacute	Random Forest	0.920	0.904	0.926	0.915	0.975	41.87
	SVM	0.949	0.940	0.952	0.946	0.989	53.05
	Logistic Regression	0.908	0.907	0.895	0.901	0.967	0.4
Acute	Random Forest	0.907	0.921	0.927	0.924	0.968	8.63
	SVM	0.939	0.963	0.937	0.950	0.987	2.52
	Logistic Regression	0.953	0.963	0.960	0.962	0.988	0.06
Subacute	Random Forest	0.933	0.932	0.919	0.925	0.992	2.35
	SVM	0.963	0.986	0.932	0.958	0.992	0.34
	Logistic Regression	0.957	0.947	0.960	0.953	0.994	0.04
Chronic	Random Forest	0.945	0.950	0.943	0.947	0.986	3.44
	SVM	0.975	0.983	0.967	0.975	0.998	0.61
	Logistic Regression	0.978	0.988	0.959	0.979	0.999	0.05

Appendix C: Models Evaluation Performance of the SPEM Dataset Summery

Table 4: Performance Evaluation of the Enhanced Data Using SPEM model, Using Different Machine Learning Classifiers, with respect to DenseNet121 Deep Learning Model

SPEM at CLAHE (Clip Limit=4)							
Stages	**Classifier**	**Accuracy**	**Precision**	**Recall**	**F1**	**AUC**	**Time(S)**
Hyperacute	Random Forest	0.9832	0.9918	0.976	0.9838	0.9995	3.30
	SVM	0.9915	0.9830	0.9945	0.9915	0.9998	0.52
	Logistic Regression	0.9957	0.9914	0.9932	0.9957	0.9999	0.04
Acute	Random Forest	0.9716	0.9534	0.9896	0.9711	0.9961	42.73
	SVM	0.9930	0.9931	0.9930	0.9930	0.9997	24.74
	Logistic Regression	0.9922	0.9919	0.9919	0.9919	0.9994	0.42
Subacute	Random Forest	0.9509	0.9398	0.9630	0.9512	0.9918	2.07
	SVM	0.9754	0.9639	0.9876	0.9756	0.9993	0.31
	Logistic Regression	0.9755	0.9753	0.9753	0.9753	0.9986	0.03
Chronic	Random Forest	0.9736	0.9649	0.9892	0.9769	0.9980	8.36
	SVM	0.9959	0.9928	0.9941	0.9964	0.9999	1.83
	Logistic Regression	0.9959	0.9928	0.9976	0.9964	0.9997	0.06