OPTIMIZING HEALTHCARE WITH MACHINE LEARNING: PROTECTING FINANCES AND IMPROVING DIAGNOSIS

Thesis Submitted
In Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY

in

Data Science

by

Vaibhavi Rajesh Mishra
(2K22/DSC/08)

Under the Supervision of
Dr. Ruchika Malhotra
(Professor, SE, DTU)

To the
Department of Software Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

May, 2024

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Vaibhavi Rajesh Mishra hereby certify that the work which is being presented in the thesis entitled Optimizing Healthcare with Machine Learning: Protecting Finances and Improving Diagnosis in partial fulfilment of the requirements for the award of the Degree of Master of Technology in Data Science, submitted in the Department of Software Engineering , Delhi Technological University is an authentic record of my own work carried out during the period from August 2022 to May 2024 under the supervision of Dr. Ruchika Malhotra.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

V. Mishra

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor (s)

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR(s)

Certified that **Vaibhavi Rajesh Mishra** (2K22/DSC/08) has carried out their search work presented in this thesis entitled **"Optimizing Healthcare with Machine Learning: Protecting Finances and Improving Diagnosis"** for the award of **Master of Technology** from Department of Software Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Place: Delhi
Date:

Dr. Ruchika Malhotra
Professor
Dept. of Software Engineering

Optimizing Healthcare with Machine Learning: Protecting Finances and Improving Diagnosis

Vaibhavi Rajesh Mishra

ABSTRACT

This thesis addresses critical challenges in healthcare through two distinct yet interrelated studies. The first paper tackles the escalating issue of financial fraud detection within healthcare systems, a pressing concern exacerbated by advancements in electronic payment methods. Traditional fraud detection approaches have proven inadequate, necessitating the development of novel solutions. This study introduces an ensemble fraud detection classifier, leveraging a combination of machine learning algorithms to enhance performance. Methodologically, the ensemble classifier undergoes rigorous evaluation utilizing accuracy, precision, and recall metrics, showcasing its superiority over conventional methods such as Naive Bayes, Random Forest, and K-Nearest Neighbours. With an accuracy of 99.46%, precision of 98.38%, and recall of 98.58%, the ensemble method significantly outperforms its counterparts, offering promising avenues for future research. Further investigations aim to integrate hybrid models tailored to address dataset imbalances and ensure real-time responsiveness in financial transactions.

The second paper addresses the urgent need for rapid and accurate diagnosis of pneumonia from chest X-ray (CXR) images, a critical aspect of medical diagnostics with profound implications for patient care. Leveraging the Swin Transformer V2, an innovative deep learning architecture, this study explores its application to pneumonia diagnosis within the medical imaging domain. Methodologically, the study evaluates the model's performance against a diverse CXR dataset, including various conditions and manifestations of pneumonia. Comparative analysis with established deep learning architectures such as AlexNet, MobileNetV3, VGG-16, ResNet 50, and DenseNet highlights the Swin Transformer V2's superiority in identifying subtle patterns indicative of pneumonia, achieving an accuracy of 98.6%. The findings underscore the transformative potential of integrating advanced deep learning models into clinical diagnostic processes, offering unprecedented accuracy and paving the way for significant advancements in healthcare practices. Possible applications of this research include the integration of advanced diagnostic models into clinical settings, potentially revolutionizing healthcare practices. Future research directions may include exploring hybrid models combining deep learning with traditional diagnostic methods and optimizing models for real-time application in clinical settings.

ACKNOWLEDGEMENT

I would like to express my deep appreciation to **Dr. Ruchika Malhotra**, Professor at the Department of Software Engineering, Delhi Technological University, for his invaluable guidance and unwavering encouragement throughout this research. Her vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan.

My heartfelt thanks go out to the esteemed faculty members of the Department of Software Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this MTech thesis.

Vaibhavi Rajesh Mishra
2K22/DSC/08

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

ML	Machine Learning
KNN	K Nearest Neighbor
NB	Naïve Bayes
RF	Random Forest
SLR	Systematic literature review
SVM	Support Vector Machine
ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
GRU	Gated Recurrent Units
SMOTE	Synthetic Minority Oversampling Technique
CXR	Chest X-ray
AUC	Area Under the Curve
CNN	Convolutional Neural Networks
DL	Deep Learning

CHAPTER 1

INTRODUCTION

1.1 Brief Overview:

The dynamic and ever-changing health care landscape presents an ongoing challenge, namely how to provide accurate and truthful medical diagnoses and how to protect financial activities from fraud. The introduction of machine learning (ML) technologies promises a new approach to these challenges. This thesis explores the dual application of ML in health care quality management by increasing the accuracy of investigations and improving the detection and prevention of financial fraud. Titled "Optimizing Healthcare with Machine Learning: Saving Finance and Improving Diagnosis," this work delves into sophisticated ML models that transform healthcare operations into more efficient, safer, and depending on patients in the integration.

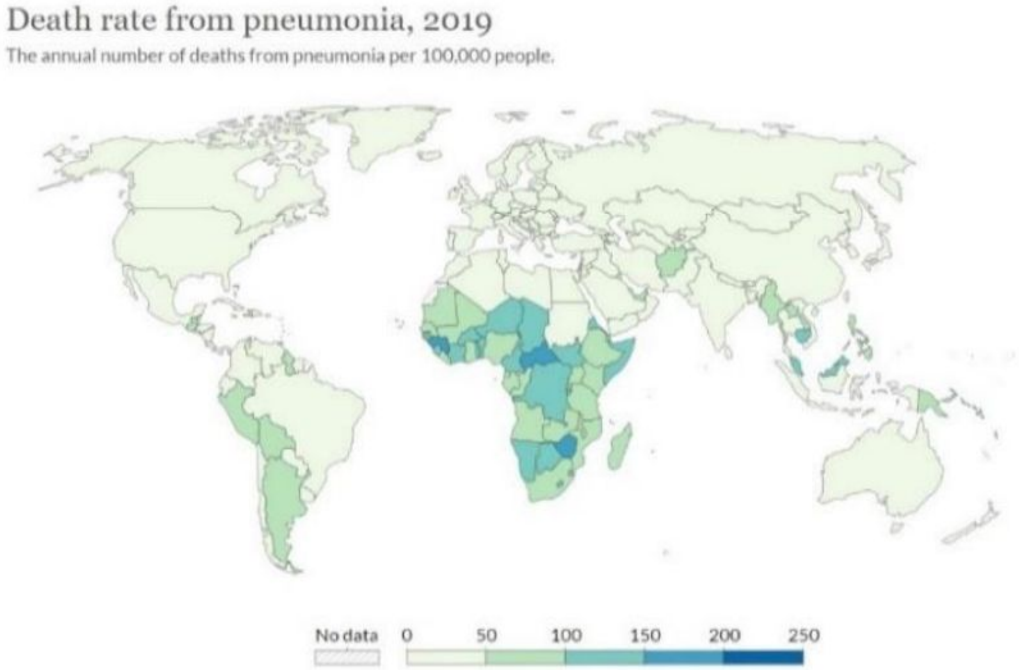


Fig. 1.1: Pneumonia Mortality rate [1]

In the world of scientific diagnostics, fast improvements in imaging era and gadget studying algorithms offer remarkable possibilities to enhance the accuracy and efficiency of clinical diagnoses. [2] Specifically, the application of the Swin Transformer V2 version to chest X-ray imaging is examined. This version represents a current method in pc imaginative and prescient, leveraging a hierarchical shape able to capturing complex patterns for more accurate detection of situations together with pneumonia. The effectiveness of this generation underscores the capacity for ML to enhance diagnostic approaches, leading to quicker and greater dependable affected person care [3].

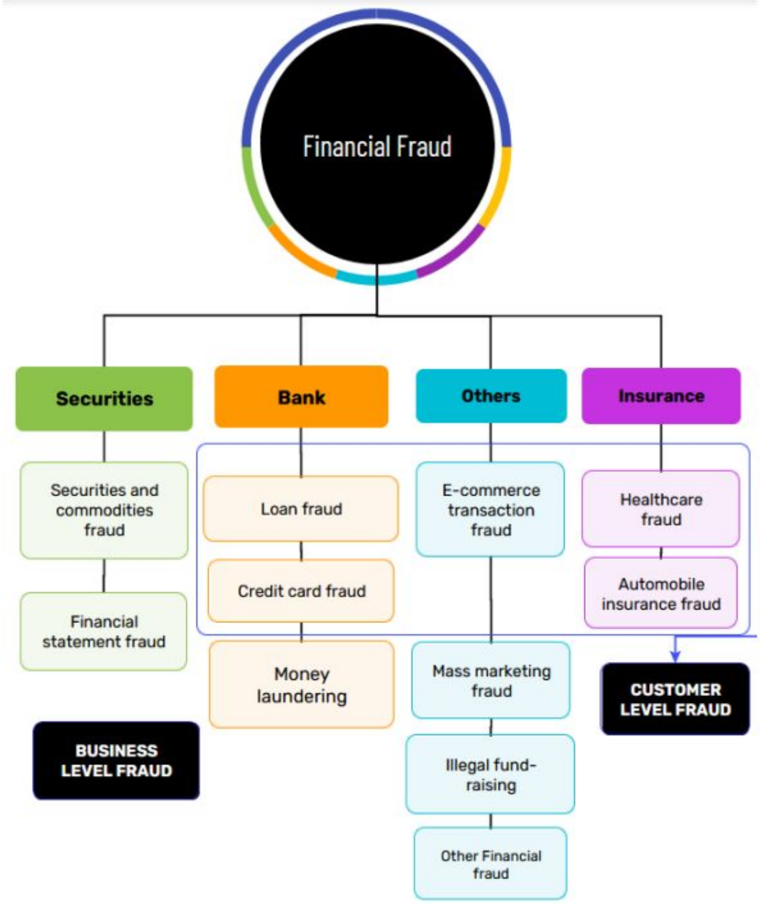


Fig. 1.2: Financial Fraud Classification

The increasing digitalization of healthcare in parallel with the rise of medical research has exposed the financial operations of healthcare organizations to a variety of vulnerabilities and fraud risks. This thesis addresses the growing need for strong fraud detection methods using machine learning algorithms, with random forests, naïve basis, and decision trees including this approach not only increases the accuracy of fraud detection but also fraud Adapts to the sophisticated strategies used, thereby protecting the financial integrity of health care systems [4, 5].

The integration of ML tools with conventional healthcare practices offers a novel way of empowering the field to tackle some persisting concerns. Thus, the possibility of applying professional ML to health providers' and employees' data processing for medical analytics and fraud detection will lead to decreasing transaction burdens while increasing the quality of patients' lives [6]. This thesis is designed to raise awareness of the immense usefulness of machine learning for healthcare applications and to increase the reader's focus on numerous advantages that could be achieved through the use of these solutions, such as making the diagnosis process easier, preventing any possible harmful influence on the patient's life, reducing the costs, and preventing financial losses by focusing on effective fraud detection techniques.

In essence, this painting seeks to demonstrate how system mastering may be strategically carried out to transform healthcare operations, making them greater efficient and resilient in the face of evolving demanding situations. Through a detailed evaluation of ML packages in diagnostic imaging and economic protection, the thesis ambitions to make a contribution precious insights and realistic answers that may be followed through healthcare companies global.

1.2 Motivation:

The reason for this thesis is some of the most crucial factors that demanding for improving the efficiency and safety of healthcare systems across the globe. In a services-based mostly society healthcare at once entails non-stop challenges in delivering first-rate care efficaciously and safely. The scope of system getting to know's (ML) implementation into healthcare operations is huge and can so help in addressing these issues mostly in diagnostic accuracy and economically sustainability.

First, motivation comes from the ongoing challenge of biomedical research. Despite advances in medical technology, diagnostic methods still hamper the accuracy and speed of diagnosis of diseases such as pneumonia, which is often dependent highly on human interpretation, Misdiagnosis and delayed treatment can have serious consequences for patient health and add to the burden on health care systems [7]. Machine learning models, especially those using advanced modeling techniques such as Swin Transformer V2, provide robust solutions. By automating and enhancing the analysis of the medical image, the models can significantly improve the accuracy and speed of diagnosis, directly benefiting the cause among patients and in the effectiveness of health care.

Second, healthcare organizations must vigorously protect their finances. As healthcare systems embrace more digital transactions, they become more susceptible to financial fraud. This yields more just than results in huge financial losses but also damages trust in healthcare institutions. The motivation has extended to exploring how machine learning can strengthen these systems against such vulnerabilities. By using sophisticated algorithms capable of detecting fraud patterns such as those found in clustering methods combining random forests, unknown bases, and decision trees, health care providers can protect their operations from the financial risks posed by fraudulent activities [8].

This thesis is driven by using the ability of system mastering technologies to revolutionize two essential components of healthcare: improving the accuracy of scientific diagnostics and improving the security of financial transactions. The twin cognizance on those areas displays a know-how of the interconnected nature of modern healthcare systems, in which improvements in a single area can drastically advantage the other. Through particular research and application of modern ML strategies, this thesis targets to demonstrate that integrating system mastering into healthcare isn't always only feasible however additionally imperative for modernizing healthcare services to satisfy modern-day challenges. By doing so, it seeks to make a contribution to the broader dreams of increasing the accessibility, reliability, and protection of healthcare for all stakeholders concerned.

1.3 Problem Formulation:

The modern healthcare industry faces the dual challenge of improving clinical outcomes through accurate diagnosis, as well as protecting against financial fraud. These goals are critical to improving patient care and ensuring the financial stability of health care providers. To address these challenges, the healthcare sector must integrate sophisticated methods for machine learning to boost diagnostic accuracy and financial security.

The primary goal of this takes a look to leverage machine getting to know to optimize healthcare transport by means of enhancing diagnostic accuracy and defensive against economic fraud. This entails developing robust fashions for the correct prognosis of sicknesses from scientific imaging facts and creating efficient algorithms for detecting and mitigating financial fraud in healthcare transactions.

The study focuses on two essential areas: medical diagnostics and financial fraud detection. In clinical diagnostics, the purpose is to decorate the accuracy and velocity of sickness prognosis the use of superior deep getting to know models applied to scientific imaging information, especially chest X-ray (CXR) photos for pneumonia detection. Accurate and speedy prognosis of pneumonia is important for powerful remedy and affected person management. Traditional diagnostic strategies frequently fall brief due to the complexity of interpreting medical photos and the subtlety of disease manifestations. Therefore, there may be a urgent need for superior version architectures, especially those based on transformers, to ensure greater green and correct diagnostics [9].

In financial fraud detection, the goal is to implement machine learning techniques to detect fraudulent activities within healthcare financial transactions. The rise of electronic payments and the increased use of credit cards have made the healthcare sector particularly vulnerable to financial fraud. Monitoring fraud committed with credit cards has proven challenging due to the sophisticated methods employed by fraudsters and the large volume of transactions that need to be scrutinized. Consequently, continuous enhancements are essential for the systems that detect fraudulent actions. Using a hybrid machine learning classifier, also referred to as an ensemble classifier, can provide a solution to this issue by combining multiple detection algorithms to increase performance.

The question being addressed in this study is multidimensional. The above case of medical diagnostics is a classic example of a task where the solution is required correctly to find sub patterns that indicate the disease in the medical images. For example, pneumonia may be expressed in mild or moderate findings in CXR that are not easy to detect using traditional methods [10]. The Swin Transformer V2 model is designed with the features that capture finer resolutions and detailed features in medical images and thus implies that the model can be employed to enhance diagnostic accuracy. This model will be trained for long periods on a large CXR data set that poses a variety of diagnostics for changes in the images of the positioning of the patient and the presence of slight manifestations of pulmonary diseases. The performance of the model will be compared with the other proposed deep learning architectures to establish the effectiveness of the approach [11].

In the case of financial fraud detection, the challenge is to create mathematical approaches, based on algorithms, that will effectively distinguish fraud from normal financial transactions. Fraudsters are always out to beat fraud detection systems by using new strategies to defraud organizations and compromise fraud protection systems. This study advocates for the use of ensemble fraud detection classifier, which works by using several machine learning algorithms to establish improved accuracy and high precision and relatively high recall. Further the result of the ensemble method will be compared with individual classifiers like RF, Naive-bay (NB), KNN. The evaluation will consist in applying the accuracy, precision and recall metrics to both classifiers.

By addressing these challenges, the study aims to make significant contributions to the body of work in medical image analysis and financial fraud detection. In medical diagnostics, the use of advanced DL models like the Swin Transformer V2 can lead to more accurate and faster diagnosis of diseases, enhancing patient outcomes and the effectiveness of healthcare. In financial fraud detection, the implementation of robust machine learning algorithms can help healthcare providers protect their finances, reduce costs associated with fraudulent activities, and enhance the overall quality of healthcare services.

In conclusion, this study seeks to optimize healthcare delivery by integrating ML techniques to improve diagnostic accuracy and financial security. The suggested algorithms and models aim to address the specific challenges in medical diagnostics and financial fraud detection, ultimately contributing to better patient care and financial stability in the healthcare sector.

1.4 Working:

This dissertation will try to explain the potential associated with machine learning technologies in revolutionizing healthcare delivery. The overall two principal objectives of this paper are: primarily, enhancing medical diagnostic accuracy and, secondarily, upgrading financial fraud detection mechanisms within healthcare systems.

It commenced by giving an extensive review of the currently available medical diagnostic and financial practices in healthcare, which brought out myriad shortcomings and unimpressive levels of efficiency. This is a background that clearly introduces how transformational machine learning can be in solving the challenges it cannot overcome. The main stages of the thesis take the high-resolution image analysis application of an advanced machine learning algorithm, the Swin Transformer V2 model. The application of this model is precisely in diagnosing pneumonia from chest X-ray images. The Swin Transformer V2 is able to discriminate between fine patterns and small variations in medical images, which often remain invisible to the human eye, with one such very advanced architecture including both hierarchical structures and shifted window mechanisms. Note that this capability is quantitatively evaluated through series on precision, recall, and the F-measure of the model against currently existing benchmarks.

Running in parallel with the diagnostic focus, the thesis explores the potential application of machine learning to detect and protect healthcare transactions from financial fraud, a truly material emerging concern accompanying the digitization of health care systems' operations. For instance, a set of machine learning techniques involving random forests, Naive Bayes, and decision trees were combined into an ensemble model to analyze complex transactional data. This ensemble approach strengthens the individual algorithm's strength in order to boost detection accuracy and adaptability to a pattern of fraud—either new or developing. Effectiveness of the approach is put to critical test through rigorous methods of validation that compare the performance vis-a-vis traditional systems for fraud detection in terms of accuracy, precision, recall, and overall efficacy.

All the methods in this thesis are data analysis-intensive, as huge datasets of both medical images and financial transactions really support them. The annotations are being carried out at an exceptionally high level of granularity in the datasets in order to guarantee very high reliability of the training and testing outcomes. Again, this further elaborates integration of ML models into existing health care systems with practicality and logistical concerns that shall lead to successful implantation. This comprises problems of data privacy and ethical pertinence, in addition to the need for continuous learning and adaptation within deployed machine learning systems to be responsive to new challenges as they come up.

In essence, this thesis illustrates a broader manner in which the power of machine learning could be further harnessed when considering optimization for health. In showing how such technologies can greatly improve both diagnostic processes and financial security measures, it not only shows the practical benefits of ML in healthcare but also charts a path toward future innovations that may further revolutionize this crucial field. Ultimately, this is intended to create a more efficient, secure, and appropriately responsive health care environment for today's globally distributed patient base.

1.5 Thesis Outline:

- The thesis discusses the transformation ability of improved diagnostic accuracy and more fraud detection within healthcare. For that, the present introduction sets the stage by elaborating, in particular, on the pivotal role that advanced machine learning technologies and Swin Transformer V2 play at large within the domain of medical imaging and ensemble methods in financial security. In other words, it goes one step further to establish the motivational factor in why machine learning should be applied: deep robust diagnostics and fraud prevention mechanisms called upon for fighting increasingly complex health data and transactions [12].
- In the literature review section, works done previously in this field and those where machine learning has been integrated into healthcare are described. It would therefore attempt to synthesize past achievements and the gaps that the current research seeks to fill, with a special focus on an application of new machine learning models that can manage more complex data sets and offer more precise predictions [12].
- The following includes important information like datasets used for training of models, specific software and library requirements, and detailed steps of the proposed methodology: data preprocessing, training of the Swin Transformer V2 model to identify pneumonia from X-ray pictures of the chest, applying ensemble techniques to detect healthcare financial anomalies in transactions, and many more.
- After the evaluation of these models, their performance would be discussed, considering their strengths and limitations interestingly. Such analysis will be

critical in appreciating how each of the models contributes toward meeting the objectives of improving diagnostic procedures alongside detecting financial fraud in varied healthcare contexts.

- It presents the outcomes with the application of the machine learning models, statistical analysis, performance metrics, and results discussed of how such outcomes can impact the overall effectiveness of enhancing health outcomes and operational efficiencies [13].
- The conclusion summarizes research results and underlines the significant improvements that have been introduced into healthcare because of machine learning. It also gives an outline for future work, suggesting how further development of research should be pursued, whether models ought to be developed further in terms of other diseases or solutions scaled up to bigger datasets. This subsequently helps in explaining the social impact, whereby machine learning will enable the social dimension to develop more patient-centric efficiency and secured healthcare systems, thus being in harmony with wider health policy goals and, thus, a boon for high-quality public health outcomes [14].

1.6 Thesis Objective

1. The objective of this project is to develop and implement state-of-the-art machine learning models that will be able to assist in diagnosing diseases with high accuracy and speed, specifically in the detection of pneumonia by using X-ray images of the chest with recent architectures available. This will be done by discovering state-of-the-art recent architectures such as the Swin Transformer V2 to use their best features of extraction capability.

2. In an effort to increase the diagnostic accuracy of pneumonia and other similar respiratory diseases, the hierarchical attention mechanism is designed with a more sophisticated Swin Transformer V2 model. The objective here will be to deal with a performance comparison against classic structures of ConvNets like AlexNet, VGG-16, ResNet-50, and DenseNet, among others.

3. Establish, curate, and preprocess a comprehensive dataset of chest X-ray images with a diversity of pulmonary conditions, patient demographics, and imaging scenarios. This dataset should properly and accurately be annotated for the purposes of training and validating machine-learning models toward robustness and generalization with regard to different clinical settings and patient populations.

4. Design and evaluation of an ensemble machine learning classifier combining detection algorithms for improving financial fraud detection accuracy, precision, and recall in healthcare transactions. This is done by making effective combinations with classifiers such as Random Forest, Naive Bayes, and K-Nearest Neighbors in order to provide better performance features in the formulation of a hybrid model for strong detection of fraud.

5. Investigate and address the specific challenges of the need for the subtle, complicated pattern of medical images to be detected in such a manner that the models put posited will be able to detect and delineate a wide variety of pulmonary pathology, including discriminating among conditions where pneumonia is part of the differential.

6. Build adaptive algorithms for financial fraud detection that adapt to the changing fraudulent tactics. Develop functions in machine learning, able to learn new patterns of fraudulent behavior and in real-time enforce adjustments to parameters in the detection process, thus reducing the false positive rate and increasing the significance of fraud-detecting systems.

7. Applying the improved diagnostic and fraud detection models in current healthcare systems in a manner that it would lift overall patient outcomes through bringing down the rate of diagnostic errors and ensuring timely and correct disease detection. Other than that, this will ensure clearly that the financial integrity of the healthcare provider is even-steven with the rest by cutting down the state of fraud healthcare services.

8. Cross-check the proposed models using a wide range of relevant metrics, including but not limited to accuracy, precision, recall, F1 score, AUC, and others, specifically on the application level. Perform comprehensive performance reviews to guarantee that developed models are, in fact, effective in the clinical and financial world, providing the benchmark for their application.

9. Evaluate the potential for scaling machine learning models to handle a large and heterogeneous dataset with high-resolution medical images. This scalability will ensure that various healthcare deployments from small clinics to major hospitals can actually leverage the potentials of these models in dealing with the volume and complexity of real-world data.

10. The findings will be presented at conferences and published in peer-reviewed scholarly publications on medical image analysis and financial fraud detection, promoting advanced machine learning techniques that advance innovation and improvements in diagnostic accuracy and financial security in the healthcare industry.

11. Transfer learning techniques can be researched and applied to the adaptation and fine-tuning of machine learning models that are intended for deployment in the various forms of medical imaging and financial datasets. This task is going to make this model more applicable in solving pneumonia diagnosis and detection problems in health care fraud and other areas of health care or any type of financial service.

12. These efforts need to include collaboration with healthcare professionals, financial experts, and technology developers so that the application solutions of machine learning can be fine-tuned to be attuned with practical needs and constraints in the health industry. This is necessary to bring these models into specific use cases, both effective and easily integrated into already-existing workflows.

13. Ethical considerations and matters on data privacy related to the application of machine learning in diagnostics within healthcare and financial fraud detection will be explored, including securely managing patient data, ensuring that machine learning models are developed and deployed while preserving patient confidentiality, and conducting them in a manner compliant with the relevant regulations.

14. Access the economic consequences of implementing advanced machine learning models in health care, including cost reduction associated with fraudulent activities, lower efficiency in diagnostic processes, and potential benefits for patient care. The evaluation shall therefore reflect a wholistic view of the value proposition for the implementation of such technologies in the health sector.

15. The updating and refinement of machine learning would remain ongoing—working in line with real-world feedback and advancements in artificial intelligence—

so that they keep gaining for healthcare diagnostics and financial fraud detection to a large extent. This will allow for continuous updating of machine learning models, along with their refinements in response to real-world feedback and advancements in the sphere of artificial intelligence, thus assuring that the models are at the cutting edge.

CHAPTER 2

LITERATURE SURVEY

Along the fight against financial fraud, the use of medical diagnostics through advanced computational techniques is quite rich and rapidly evolving. In finance, earlier efforts in fraud detection resorted to rule-based systems and statical approaches; although important, they are simply not sufficient when trying to cope with sophisticated and adaptive fraudulent behaviors. As machine learning was rising, research started into algorithms like Naive Bayes, Random Forest, and K-Nearest Neighbors, which could provide a little better performance in detection. However, these largely failed to accommodate the scale and complexities of data pertinent to real-world scenarios. More recently, a follow-up has been the ensemble methods, which combine multiple algorithms; harness their combined strengths; offer large improvements in accuracy across the metrics of detection, precision, or recall; and go the extra mile to control for overfitting. On the other hand, medical diagnostics has undergone a paradigm shift with the application and development of deep learning. Besides, new promising enrichment in this area with the advent of CNNs in the domain of conventional techniques for image analysis, accompanied or guided by manual interpretations from radiologists, was followed by the introduction of transformer-based models. The Swin Transformer V2, which comes with a new self-attention mechanism, shift-window based, reports state-of-the-art performance for the pneumonia identification from chest X-rays task, outperforming previous architectures: AlexNet, MobileNetV3, VGG-16, ResNet 50, and DenseNet. This literature research does indeed strive to delve deeper into such developments by looking at the historical context and current innovations in the applications of machine learning and deep learning in the arena of healthcare.

Ali et al., (2022) [15] undertook a systematic literature review (SLR) to consolidate studies focusing on ML-driven fraud detection. They adopted the Kitchenham method to methodically identify, extract, and collate relevant articles, ensuring a structured synthesis of findings. This review sourced studies from notable electronic database libraries, emphasizing ML techniques for fraud detection, prevalent fraud types, and evaluation metrics. From the analysis, it became evident that SVM and ANN stand out as primary ML methods in fraud detection, with credit card fraud as the predominant concern addressed through ML. The study culminates by spotlighting challenges, gaps, and constraints in financial fraud detection, suggesting avenues for future research.

Amponsah et al., (2022) [16] suggested that identification and prevention of healthcare fraud are possible using blockchain technology and machine learning methods, particularly during the claims processing phase. Decision trees classify the initial claims dataset. Subsequently, the gathered data is saved in an Ethereum smart contract, tailored for identifying and mitigating healthcare fraud. Experiments comparing different tools demonstrate that the top performer has a sensitivity of

98.09% and an accuracy of 97.96% for classification. As a result, the suggested solution improves the 97.96% accuracy with which blockchain smart contracts can identify fraud.

Yıldırım et al., (2022) [17] examined that using machine learning strategies identification of financial fraud could be scrutinized in depth. This study focuses on three distinct forms of fraud within the finance sector: insurance fraud, corporation fraud, and bank fraud. Most machine learning-based studies were carried out in the banking fraud industry, as shown by this review's conclusions that SVM, ANN, Decision Trees, and Random Forests are the most often used methods for financial fraud detection. This study demonstrates that deep learning and machine learning applications based on ensembles have been more popular over the last few years to improve the effectiveness of financial fraud detection.

Turaba et al. (2022) [18] highlighted the prevalent challenge of detecting online credit card fraud. Such fraud can arise when culprits use purloined cards for unauthorized transactions or when a fraudster misuses credit card information. Addressing these issues necessitates an efficient system for recognizing credit card fraud. Current methods leverage both machine learning and deep learning for expedited and effective solutions. This research delves into the examination and outcomes of detecting fake credit card activities. This study evaluates various methods including Adaptive Boosting, CNN, and a fusion of CNN with GRU. To counter the dataset's skewed distribution, the Synthetic Minority Oversampling Technique is employed. Of all the techniques assessed, the Convolutional Neural Network exhibited the best performance in metrics such as AUC-ROC, accuracy, precision, and recall, showing enhanced precision compared to earlier studies.

Mehbodniya et al., 2021 [19] also point to the tremendous corpus of health and financial data collected by this industry. The increased use of electronic payments is causing it to be difficult and expensive in the monitoring of credit card fraud among the healthcare service providers. Many and diversified conventional transaction information data were used among such datasets to train machine learning as well as deep learning classifiers, namely K-Nearest Neighbor (KNN), Random Forest (RF), Naive Bayes (NB), Sequential Convolutional Neural Networks (SCNN), and Logistic Regression (LR). The models were tested on publicly available datasets, resulting in accuracies of 96.1% in NB, 94.8% in LR, 95.89% for KNN, 97.58% in RF, and 92.3% for SCNN. Of those, KNN has been the best performing.

Asha et al., (2021) [20] analyzed that Nowadays, Overuse of payment by credit card options increases the likelihood of credit card fraud. This is due to the rise in purchasing via the internet and the corresponding rise in fraud cases, which in turn results in significant financial losses. Since this is the case, finding efficient means to cut down on it is essential. Fraudsters also use masquerade assaults, phishing schemes, and other similar techniques to acquire users' credit card details. The purpose of this study is to use many Machines Learning methods, including the SVM, KNN, and ANN, to forecast the incidence of fraud. Furthermore, the author separates the successful deep learning supervised and machine learning strategies used to distinguish fraudulent from legitimate transactions.

Azhan et al., (2020) [21] introduced that fraud is any action intended to harm someone financially. As digital money grows in numerous nations, so do digital money scams. Credit card firms and banks lose billions to such theft every year, affecting their income and workers. This study considers credit card theft and explains how Neural Networks and Machine Learning could identify prospective fraudsters based on their past failures and information. Logistic Regression, Support Vector Machine, Multinomial Naive Bayes, Random Forest Regression, and basic Neural Networks are used for Machine Learning.

Anum Masood et al. (2023) [22] have evaluated the performance of new deep learning techniques on medical image analysis datasets, including ILD, ANODE09, LIDC-IDRI, ELCAP, and LUNA16. They used an ensemble from ST-MSMLFFR, Swin Transformer, and Unet to achieve 96% Dice Similarity Coefficient. It describes a way that very high accuracy can be reached in medical image segmentation and yet provides the potential to enhance these models for improvement of diagnostic precision in highly complex imaging scenarios.

Xinli Wu et al. (2023) [23] experimented with Xray Dataset and ChestXray dataset from Kaggle by applying multiscale Swin Transformer to better analyze the image. Multiscale processing has been proven pretty effective for capturing fine features in the chest X-ray images because the performance of the presented model shows an average accuracy of about 91%. This work documents how the transformer-based models have revolutionized the discipline of medical diagnostics.

Anis Amirah Binti Ramli et al. (2023) [24] applied a Convolutional Neural Network by using the CXR Dataset for classification of any medical condition reported on a chest X-ray. The CNN model is highly accurate, at around 96%, indicating that the technique is robust and reliable for tasks related to the classification of medical images. This present work again confirmed the application of CNNs in medical imaging.

Zixun Ye (2022) [25] The Swin Transformer was validated by Zixun Ye using text-free tasks on datasets widely utilizing, including CASIA-WebFace, MS1M-ArcFace, and LFW Databases. Managed to get 60.6% in total model parameter reduction, marking huge gains in computational efficiency without a dent to performance. This is another great example of how well-performing and computationally efficient the Swin Transformer is for other areas besides medical imaging.

Mohammad Yaseliani et al. (2022) [26] used a combination of the Support Vector Machine, Radial Basis Function, and Logistic Regression algorithms in one ensemble model to analyze the CXR dataset. This model returned an impressive 98.55% accuracy index, demonstrating that ensemble approaches can harness the power of multiple algorithms to boost diagnostic accuracy methods in medical imaging.

Fangfang Li (2022) [27] proposed the use of an SPD module combined with a Focal Transformer for applying medical images to the Lung-Xray Dataset. The results indicated an experimental accuracy of 95% and showed a good capability for solving complex cases of medical image processes. This work had exhibited the key benefit of transforming the specialized modules into transformer architecture in order to achieve better performance.

Md. Nahiduzzaman et al. (2021) [28] A hybrid model of CNN, PCA, and ELM combined with the CXR Kaggle pulmonary dataset showed that images with this model have accuracy in dimensionality reduction, with machine learning holding potential for enhanced diagnostic accuracy of 98.32%.

Jianpeng Zhang (2021) [29] worked on the CAAD Model from X-VIRAL provided a sensitivity of 71.70%, with the Area Under the Curve (AUC) being at 83.61%. The model's sensitivity being moderate, the AUC represents the overall capacity to discriminate over different classes. The current study shows challenges and potentials in the development of models for viral infection detection through the analysis of chest X-rays.

Harsh Sharma (2020) [30] The accuracy in the Convolutional Neural Network applied to the CXR Dataset was 90%. This has indicated that although the results are less when compared to some of the other studies, this work has made a clear statement of proof toward the continued relevance and application of CNNs in medical diagnostics, hence laying the firm foundation toward improvements further in accuracy and robustness.

José Raniery Ferreira Junior (2020) [31] developed a Multi-view Ensemble CNN with analysis of the CXR Dataset and reached an accuracy of 92%, meaning that with many perspectives and ensemble techniques, it is possible to advance the disorder diagnostic accuracy from medical images.

Gaurav Labhane (2020) [32] they achieve an accuracy well above 97% for their CXR dataset using a mix of CNN and transfer learning. This high accuracy is, once more, due to a beneficial aspect of transfer learning where pre-trained models on large datasets are fine-tuned with domain-specific data to have a much-improved performance on medical image analysis.

TABLE 2.1: COMPARISON OF RELATED WORK

Author	**Dataset**	**Technique**	**Outcome**
Ali et al., (2022) [15]	Finance dataset	SVM and ANN	Key difficulties, unanswered questions, and theoretical boundaries in financial fraud detection are outlined, along with potential directions for further study.
Amponsah et al., (2022) [16]	Kaggle dataset	Blockchain Technology and ML Methods	The proposed fix increases the 97.96% success rate of fraud detection via blockchain smart contracts.
Yıldırım et al., (2022) [17]	Kaggle	Machine Learning	Ensemble-based ML & DL applications have gained attraction in the fight against financial crime.
Turaba et al., (2022) [18]	Private	Machine Learning	This finding is better than prior work in terms of accuracy.
Mehbodniya et al., (2021) [19]	Public	SMOTE	KNN outperformed other methods in the comparison analysis.

Asha et al., (2021) [20]	Finance dataset	Supervised ML and Deep Learning	The study distinguishes effective supervised machine learning and deep learning algorithms for identifying fraudulent transactions.
Azhan et al., (2020) [21]	Private	Neural Networks and Machine Learning	This study shows how Machine Learning and Neural Networks could detect credit card thieves based on prior failures and information.
Anum Masood., (2023) [22]	ILD, ANODE09, LIDC-IDRI, ELCAP and LUNA16	ST-MSMLFFR, Swin Transformer, Unet	Achieved 96% DSC
Xinli Wu., (2023) [23]	Xray Dataset and Chestxray Dataset (Kaggle)	Multiscale Swin Tansformer	The model achieved the average accuracy of approx. 91%
Anis Amirah Binti Ramli., (2023) [24]	CXR Dataset	CNN	The achieved accuracy is 96%
Zixun Ye., (2022) [25]	CASIA-WebFace, MS1M-ArcFace, LFW	Swin Transformer	Reduces the total amount of model parameters by 60.6%
Mohammad Yaseliani., (2022) [26]	CXR Dataset	Ensemble (SVM + RBF +LR)	The achieved accuracy is 98.55%
Fangfang Li., (2022) [27]	Lung-Xray Dataset	Focal Transformer integrated with SPD module	The achieved accuracy is 95%
Md. Nahiduzzaman., (2021) [28]	Kaggle CXR images	CNN-PCA-ELM	The achieved accuracy is 98.32%

Jianpeng Zhang., (2021) [29]	X-VIRAL dataset	CAAD Model	Reaches a sensitivity of 71.70% and an AUC of 83.61%.
Harsh Sharma., (2020) [30]	CXR Dataset	CNN	The achieved accuracy is 90%
Jos´e Raniery Ferreira Junior., (2020) [31]	CXR Dataset	Multi view Ensemble CNN	The achieved accuracy is 92%
Gaurav Labhane., (2020) [32]	CXR Dataset	CNN and Transfer Learning	The achieved accuracy is over 97%

CHAPTER 3

METHODOLOGY

3.1 Dataset Description:

This will render advanced machine learning techniques even more invaluable, as this dataset will help achieve twin objectives: optimizing healthcare diagnostics and financial integrity. The large dataset had been carefully gathered to consist of the widest medical and financial information crucial in embedding strong individual machine learning models for optimum disease diagnosis and sniffing out any fraudulent activities.

The medical diagnostics part mainly contains chest X-ray (CXR) images regarding the diagnosis of pneumonia. Pneumonia is a common respiratory condition, usually caused by infections and that easily becomes lethal; it is an important diagnostic dilemma because its manifestations in CXR tend to be subtle and varied. The dataset includes thousands of images at very high resolutions pooled from a variety of sources, ensuring wide demographic representation of patients going through different states of image acquisition conditions and pulmonary states. This is important because this variation becomes actually necessary in the process of training a machine learning model to identify the idiosyncrasies in the patterns that develop for pneumonia and other lung ailment [33].

Fig. 3.1: CXR Dataset [33]

The dataset comes along with very elaborate annotations for each image conducted by MDs, who thoroughly note whether the pneumonia is of bacterial, viral, or COVID-19 type. Other clinical information can be derived from the annotations, which include age, gender, and medical histories. This rich metadata provides machine learning models with the much-needed diversity of examples; it fuels generalization across different patient populations and clinical scenarios.

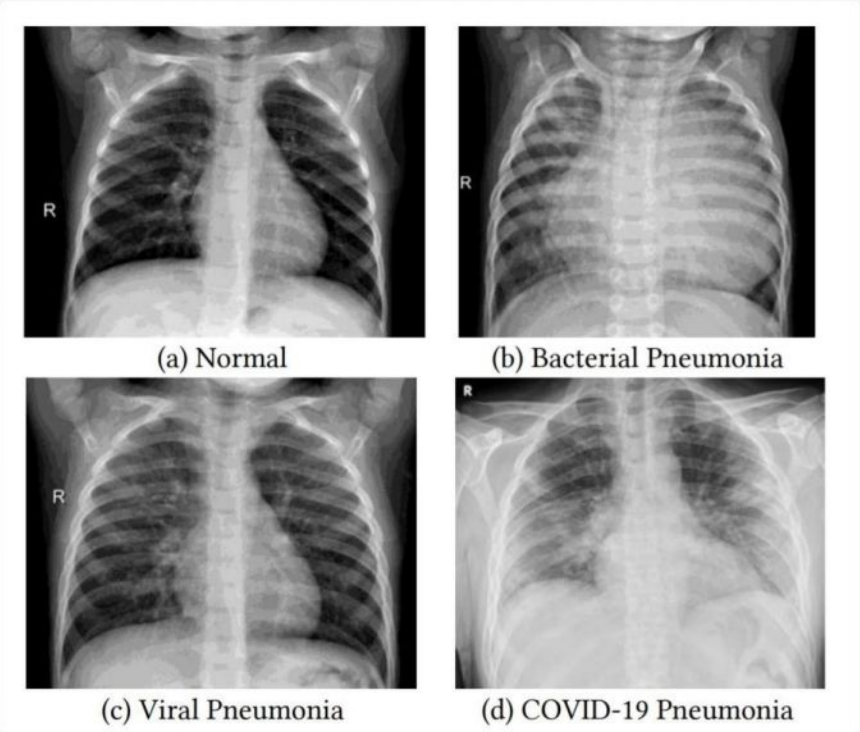


Fig. 3.2: Types of Pneumonia [34]

This adds even more generalization capacity to the diagnostic model, including tricky cases with low image quality or patient positioning variability, as well as overlapping symptoms with other respiratory pathologies. Such diversity warrants that model accuracy is not only ideal but robust against real-world variation in medical imaging.

The set of data used here is a financial transaction record in the healthcare sector, which is large-scale and involves millions of related electronic payments, credit card uses, insurance claims, and other health service activities. The record of transaction will have many attributes per detail: the amount of the transaction, the accounts of origin and destination, timestamps, and the type of transactions—payment, cash-in, cash-out, transfer, or debit.

This dataset is also labelled based on historical data and verification by experts, whether the transaction is genuine or fraudulent. Fraud in the banking system manifests through a number of routes: identity thefts, false claims on insurances, illicitly gaining access to a person's financial accounts, among others. This is critical for training machine learning models to distinguish between normal and suspicious activities.

Fraud detection—general or financial—is an issue properly addressed only after the dataset has been greatly enhanced by domain-specific knowledge, advanced feature engineering techniques, and most of these features. In fact, all of these features allow models to hunt for tiny patterns and anomalies that give evidence of the behavior being illegitimate. Other features derived from a data set include transaction frequency, average transaction amount, deviation in habitual spending patterns, and relations between multiple transactions.

Besides that, the fraud patterns of the dataset and new upcoming threats of the health business have to be updated regularly. This updating process ensures that the machine learning models work effectively against all new means of fraudulent tricks, being updated to give reliable protection for the associated finance in taking care of healthcare.

Overall, the dataset for this current study represented a complete and finely curated set of medical and financial data for utilization in crafting advanced machine learning models, directed towards the optimization of healthcare diagnostics and financial integrity. Diverse and detailed examples, challenging cases, and records that are always up-to-date mean that in the end, only the models that really burn the midnight oil in high real-world healthcare and financial environments will succeed—to reach high accuracy, robustness, and adaptability. This general approach to the collection and preparation of data would be quite essential for success in application so that machine learning solutions may ensure that better patient care is taken, while health resources are protected.

The following table provides the details of the dataset for optimizing healthcare with machine learning, with attention to safeguarding finances and intensifying diagnostics:

TABLE 3.1: Dataset Description

Dataset Component	**Description**	**Attributes**	**Purpose**
Chest X-Ray (CXR) Images	High-resolution images of chest X-rays collected from diverse sources, focusing on pneumonia detection.	Image ID, Resolution, Patient Age, Gender, Diagnosis (Pneumonia type: bacterial, viral, COVID-19), Image Quality Indicators, Annotation by medical experts.	To train and validate machine learning models for accurate and rapid diagnosis of pneumonia and other lung conditions.
Medical Annotations	Detailed labels and clinical information associated with each CXR image.	Pneumonia Presence (Yes/No), Pneumonia Type, Severity, Additional Clinical Notes.	To provide rich metadata for machine learning models to learn from diverse and accurately

			labeled medical images.
Patient Demographics	Information about the patients whose CXR images are included in the dataset.	Patient ID, Age, Gender, Medical History, Smoking Status, Geographic Location.	To ensure that the models generalize well across different patient populations and clinical scenarios.
Diagnostic Challenges	Images presenting diagnostic challenges such as poor quality, variations in patient positioning, and overlapping symptoms with other conditions.	Image Quality Score, Positioning Notes, Overlapping Conditions, Repeated Imaging.	To make models resilient to real-world variability in medical imaging.
Financial Transactions	Records of financial transactions related to healthcare services, including electronic payments and credit card usage.	Transaction ID, Timestamp, Transaction Amount, Origin Account, Destination Account, Transaction Type (Payment, Cash-In, Cash-Out, Transfer, Debit), Fraud Label (Yes/No).	To train and validate machine learning models for detecting and preventing financial fraud in healthcare transactions.
Transaction Attributes	Detailed features of each transaction to capture patterns indicative of fraud.	Frequency of Transactions, Average Transaction Amount, Transaction Deviation, Account Balance Changes, Relationships Between Transactions.	To enhance the models' ability to detect subtle anomalies and fraudulent behavior.

Fraud Labels	Historical data and expert verification indicating whether each transaction is legitimate or fraudulent.	Fraudulent (Yes/No), Fraud Type (Identity Theft, False Claims, Unauthorized Access).	To provide accurate labels for supervised learning and improve fraud detection accuracy.
Derived Features	Features engineered from raw transaction data using domain-specific knowledge.	Transaction Patterns, Time-Based Features (e.g., Peak Hours), Geographic Patterns, Anomaly Scores.	To improve the models' ability to capture complex fraud patterns and reduce false positives.
Updates and Maintenance	Regular updates to include new fraud patterns and emerging threats in the healthcare sector.	Update Timestamp, New Fraud Patterns, Emerging Threats, Data Corrections.	To ensure that the models remain effective and adaptive to evolving fraudulent tactics.
Evaluation Metrics	Metrics used to evaluate the performance of the machine learning models on both diagnostic and financial fraud detection tasks.	Accuracy, Precision, Recall, F1 Score, Area Under the Curve (AUC).	To assess the models' efficacy in real-world clinical and financial environments.

This table enables an integrated overview of the dataset, its components, and their purposes in this framework for optimum repositioning of health through machine learning. By doing so, very much embodies research in its objective of increasing diagnostic accuracy and financial security within the healthcare industry.

3.2 Evaluation Parameter:

1. **Accuracy**: A performance metric called accuracy gives a broad view of how accurate a model is. It is quantified as the percentage of accurate forecasts (including true positives and negatives) for all events in the data. Although this fact is easy to interpret and is widely used, it may not be appropriate for unequal classes of data where one class is more abundant than the other.

$$Accuracy = \frac{True\ Positives\ +\ True\ Negatives}{Total\ Instances} \tag{1}$$

2. **Precision**: A metric called precision evaluates how well the model can identify the model that is expected to perform well. The ratio of real positives to the total of real positives and negatives is used to compute it. Precision is especially important when the false positive rate is high because it focuses on the accuracy of the prediction quality. High accuracy values demonstrate how well the model prevents erroneous positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Postives} \tag{2}$$

3. **Recall**: Recall evaluates how well the model recognizes every instance of positivity in the data. The ratio of real positives to the total of false positives and real positives is used to calculate it. In cases where missing values lead to significant results, regression becomes important as it demonstrates the model's ability to avoid negative values. The model's ability to identify the majority of beneficial occurrences in the data is suggested by a high recovery rate.

$$Recall = \frac{True\ Positives}{True\ Positive + False\ Negatives} \tag{3}$$

4. F1 Score: The F1 score acts as a balanced average of precision and recall. It evaluates a model's ability to accurately identify positive instances while minimizing both false positives and false negatives.

$$F1\ Score = \frac{2 * (Precision * Recall)}{(Precison + Recall)} \tag{4}$$

3.3 Proposed Method:

The proposed methodology for this thesis is exclusively based upon new, latest state-of-the-art machine learning techniques in order to solve two of the most prominent problems: improvement in the accuracy of medical diagnoses using imaging and improvement in the detection of financial fraud within healthcare systems. It is a combination of theoretical frameworks with practical applications, drawing from the latest developments in research and technology on machine learning.

This study has used the diagnostic component, the Swin Transformer V2, because it is superior in the task of processing high-resolution images. All these models are first-rate and find good use for medical imaging tasks like detecting pneumonia in chest X-ray images. This may be better than regular convolutional neural networks at capturing complex patterns and subtle nuances in medical images through hierarchical structures and the use of shifted window mechanisms. The methodology includes a first step: normalization and quality enhancement of X-ray images during the pre-processing phase; then, the images will pass through a segmentation process into patches that are fed to the Swin Transformer. The process is designed to construct the model in an optimized way such that it learns from the diversity of pathological features present in the images.

At the same time, the financial fraud detection component in this dissertation uses a machine learning ensemble and employs three robust algorithms: Random Forest, Naïve Bayes, and Decision Trees. The ensemble method is adopted to exploit mutually exclusive strengths of the individual algorithms and bring the overall effectiveness and accuracy of the fraud detection process to the next level. This component, therefore, involves collecting and preprocessing rich transactional data from healthcare systems in which multiple features, such as the amount of a transaction, details of an account, type of transaction, and so on, can be extracted to train the ensemble model. Great attention to feature engineering can reveal anomalous patterns characteristic of fraud.

The models will then be optimized in the two aspects of the study and be cross-validated to find the best parameters so as to prevent overfitting, thereby building strong and generalizable models across different data sets. Performances are systematically recorded and analysed for metrics like accuracy, precision, recall sensitivity, F1-score, among others.

Equally important is to look into how these machine learning algorithms would eventually be integrated within an operational workflow in healthcare. In simple terms, this means developing interfaces such that the models developed can be made to communicate with already existing healthcare information systems such that in a real-world setting, the models can function effectively in devolved usage. Other emerging issues regarding scalability, data privacy, and security are also used to develop protocols to deal with data in an ethical manner while in strict compliance with relevant standards of regulatory requirements.

In summary, the proposed methodology aims to not only prove the effectiveness of using machine learning models like the Swin Transformer V2 with ensemble algorithms to enhance health diagnostics and financial security but also to develop a blueprint for implementing these advanced technologies into day-to-day operations within the respective health sector.

Here is a detailed table which presents the proposed methodology in the paper titled "Optimizing Healthcare with Machine Learning: Protecting Finances and Improving Diagnosis":

TABLE 3.2: Proposed Method

Component	Description
Diagnostic Imaging	**Model Used:** Swin Transformer V2 **Objective:** Enhance the approximate rate of diagnosis of pneumonia using chest X-ray images. **Methodology:** 1. **Pre-process Images:** Normalize images and improve image quality before analysis. 2. **Image Segmentation:** The process of segmenting an image and decomposing it into input patches for the Swin Transformer. 3. **Model training:** The availability of hierarchical structure and shifted window mechanisms is enabling the possibility to capture more detailed features. 4. **Validation:** Use cross-validation techniques for parameter optimization and preventing overfitting.

Financial Fraud Detection	**Models Used:** Ensemble of Random Forest, Naive Bayes, and Decision Trees **Objective:** To increase the probability of identifying fraud in health financial transactions. **Methodology:** 1. **Data Collection and Preprocessing:** Collect sufficient transactional data and extract the primary features from them; a transaction's amount, account information, or transaction type. 2. **Feature Engineering:** Develop features to let an algorithm detect some out-of-the-ordinary patterns that may be fraudulent. 3. **Model Training:** Ensemble multiple algorithms to balance strengths and weaknesses. 4. **Validation:** Measure performance using metrics based on accuracy, precision, recall, and F1 score.
Integration and Implementation	**Goal:** To ensure that the machine learning models are effectively integrated into the operations workflow of healthcare. **Methodology:** 1. **System Integration:** Design applications along with linking to existing systems of information in healthcare. 2. **Scalability:** Address the issues in scalability so that the models can be used effectively for complex, large-scale data. 3. **Data privacy and security**: Develop protocols for handling data in an ethical and regulatory manner. 4. **Real-World Testing:** Apply the model to real-life problems and prove its effectiveness and usefulness.

3.4 Necessary Libraries:

- NumPy and Pandas are libraries commonly used for data manipulation and analysis. For the data fed into the models, these libraries have a set of tools that cater to dealing with large sets of data and are very beneficial.
- Scikit-learn: for easily and effectively implementing traditional machine learning algorithms. It is most often used in the training of the model, cross-validation, and calculation of a majority of machine learning metrics.
- Build and train more complex neural network architectures: TensorFlow and Keras—these libraries are maintaining the models for all deep learning techniques such as LSTM and Convolutional Neural Networks (CNNs).
- PyTorch can also be used for dynamic computation graphs and memory efficiency, which is useful in cases where the model deals with relatively large datasets.
- Matplotlib and Seaborn are used for creating static, interactive, and animated visualizations to analyze the model outputs and data insights effectively.

3.5 Software Requirements:

The set-up drivers require software that is able to implement and test created machine learning models. This rugged software has to be capable of conducting bulky data processing, model training and validation, along with scenario simulation inside, prior to the installation of these models.

- This is what makes Python the most important programming language: it has rich support in data analysis and machine learning, supported by libraries and frameworks best suited for this functionality. Due to its simplicity and flexibility, Python is best in taking up lots of demands from machine learning projects pertaining to systems.
- A Jupyter Notebook is an interface in which you can code, debug, and visualize data and results. Anyway, all these within an interactive environment that makes it excellent for iterative tests and model tuning.
- Git was used to handle version control, to ensure that the research is reproducible
- Simulation tools, such as SUMO (Simulation of Urban Mobility) or VISSIM, may be applied to model scenarios and examine the effects of proposed machine learning interventions that the models suggest under a simulated circumstance. Throughout this document, tools such as these will be referred to in parallel with some other tools that make it possible to investigate how models would work if applied in practice
- High-performance computing resources and cloud computing platforms from AWS and Google Cloud are used to handle all the data processing of massive scale and model training with high complexity, otherwise requiring big computational power

3.6 Data Preparation:
This part of the thesis explains these two steps: data cleaning and preprocessing to be carried out correctly in the application of machine learning in this thesis in order to realize better results. Since data quality is at the very essence of ensuring value from machine learning models, these initial stages are very important in guaranteeing the validity when it comes to the results and predictions made.

3.6.1 Data Cleaning:
The first step in the preprocessing series of activities involves the careful process of cleaning data. Some of the steps include:

- Inconsistencies Removal: Identify and remedy time inconsistencies in data collection; correct sensor malfunctions and misalignments with data from other sources.
- Imputation of missing values: This presents a great imputation technique to use for selected missing value fields. The choice between the different methods of Imputation (average, interpolation, etc.) is a function of data characteristics and the expected model results that it will affect.
- Outlier Filtering: It uses statistical techniques to spot those extremities that are undetectable and, hence, exclude such observations from the analysis, which could alternatively influence the findings. This is especially relevant in the data of medical because the records of any anomalies in the usual medical patterns do not fit appropriately, consequently damaging the training of the models.

3.6.2 Data Preprocessing:

Data cleaning has been done; now is data preparation that involves optimizing the machine learning models. This involves a couple of key processes:

- Feature Selection: Identifying the most relevant features that are likely to contribute to improved medical and safety. It may include data points such as Claim Pattern, Payment mode, Type of pneumonia, etc.
- Feature engineering: create new features out of existing data, which might further help a model learn. For example, turning timestamps into categorical features like 'time of day' or 'day of the week'.
- Normalization and scaling: Standardize the feature range, but be careful about dominance and scale of a feature. In some approaches, the Min-Max scaling of the Z-Score normalization is acceptable.
- Data preprocessing: This is to modify all data into machine-learning-model-friendly formats. I will reshape data into sequential windows for time series and LSTM networks.
- Data augmentation is simply the process of artificially expanding the available data diversity so as to get more examples during training. It offers greater robustness of models in cases of tasks like image recognition through processes such as rotation, scaling, and cropping.

These are all preprocessing steps in which the sources of potential errors in training the models decrease, and contrastingly, their predictive performance increases. It is to ensure that the data going into the models is clean, relevant, and structured in a way that suffices for the analytical objectives set forth.

CHAPTER 4

DEEP LEARNING AND ENSEMBLE

In this paper, "Optimizing Healthcare with Machine Learning: Protecting Finances and Improving Diagnosis," we apply the latest state-of-the-art machine learning techniques to achieve the following two aims for health care: improved diagnose accuracy and protected financial integrity. Appropriate techniques are rigorously chosen that form a solid foundation for generalizability, accuracy, and adaptability in real-world scenarios.

We used the Swin Transformer V2 model. It is quite an advanced deep learning architecture which showed much better results for image analysis-related tasks compared with all the others. In its turn, a Swin Transformer V2 applies a hierarchical vision transformer structure with shifted windows. This makes it efficient in capturing complex patterns in medical images. Such kinds of models have also shown very good performance on high-resolution chest X-ray images, where subtle manifestations of pneumonia have been identified with high precision. In this type of technique, the images are split into non-overlapping patches that use self-attention mechanisms to get detail and accuracy in the radiograph.

Very diverse datasets with a large collection of CXR images combined with clinical information and annotated with detailed information on the presence and type of pneumonia were effectively used during training. The model serves to be more robust and capable of better performance, considering the assortments of imaging conditions and patient demographics. Data augmentation techniques, with respect to rotations, flips, and scaling, had been applied during training to make the model strong in performance. It can also be in the form of transfer learning, first training a model over a large dataset and later fine-tuning on our specific CXR dataset for higher diagnostic performance.

Detection of financial fraud in health care was addressed using a combination of the more traditional and ensemble machine learning classifiers. The very nature of healthcare financial fraud is complex and changing; hence, it is very suitable for the ensemble method. In this research, we combined several existing techniques, including Random Forest, Naive Bayes, and K-Nearest Neighbor Classifier, to develop a hybrid model constraining the benefit of individual classifiers. This ensemble approach ensures better overall detection with low chances of false positives and improves the model's ability to detect very subtle fraudulent patterns.

The training dataset for the financial transaction fraud detection model will consist of millions or more, where each will be properly labeled based on historical data and expert verification as real or fake. Several features to be taken into account in every transaction contain the amount, origin and destination of accounts, type of transaction, and timestamps. Features are derived in terms of transaction frequency, deviation from normal spending patterns, and the relationship between multiple transactions. These features improved the model in detecting anomalies that may be indicative of fraud.

These data are partitioned into a training and validation set to train the ensemble classifier. The optimized model parameters and model performance metrics, which include accuracy, precision, recall, and F1 score, are obtained based on the classification of test cases using cross-validation methods. Then we test the ensemble model using a validation set that represents real-world transaction scenarios to ensure its effectiveness in fraud detection and prevention during financial health care transactions.

We implement continuous learning mechanisms to keep the model up-to-date and effective in new patterns of fraud corresponding to emergent threats. This is made possible by updating transaction records in the dataset and frequent retraining of the model to detect evolving fraudulent behavior.

In a nutshell, the methods take into account the current state-of-the-art in machine learning models and apply ensemble toward optimally addressing healthcare diagnostics and financial fraud detection. We use the Swin Transformer V2 for medical image analysis and an ensemble of classifiers for fraud detection to come up with high-accuracy, robust, and adaptable solution approaches. This integrated system, therefore, does well not only to enhance attention to patients through proper and timely diagnosis and treatment but also to protect the financial resources of health providers by ensuring that any incidence of fraud is properly detected and brought down.

4.1 Swin Transformer V2 for Medical Diagnostics

Swin Transformer V2 has a hierarchical design, and it uses shifted windows in its self-attention mechanism. This enables the model to capture both local and global features from an image in order to do a better, deep analysis. In which a manner to structure the input image hierarchically such that it converts into nonoverlapping patches with its processing distributed among several levels of hierarchy, abstracting gradually from local to global features within an image. This is critical in medical imaging, as the opportunity to detect some subtle patterns or anomalies may be large to make a proper diagnosis [35].

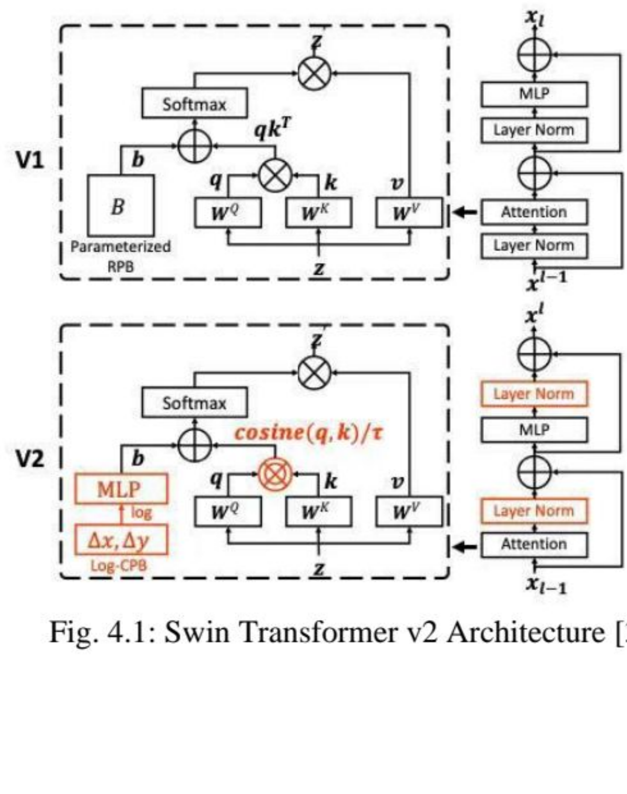


Fig. 4.1: Swin Transformer v2 Architecture [36]

The conceptual breakthrough of modeling with global self-attention in the Swin Transformer V2 mainly further advanced the powerful progress driven by mathematical advances. In vision transformers prior to this, a model using global self-attention partitioned features into windows. Self-attention computation was then performed within each window. Afterwards, these windows were shifted between the layers so that the model captured cross-window dependencies/interactions. Clearly, this would help the model gain a better understanding of the spatial relationship of an image and subsequently identify the important details in a medical scan that usually contains many intricate details [37].

Another important characteristic of the model is its ability to work efficiently with high-resolution images; normally, medical images, such as CXRs, happen to be high resolution and full of much information that cannot be compromised at the time of making conclusive diagnoses. The design allows the Swin Transformer V2 to handle these high-resolution images without incurring overly high computational costs; therefore, it remains very powerful and practical for use in real-world applications within medical environments.

In the same way, below, we train the Swin Transformer V2. We first source a very rich dataset containing CXR images from various cases like pneumonia in different strains and severity. This dataset is then meticulously annotated under the interpretation of medical professionals, providing the desired fine-grained labels to guide the model during training. Augmentation is done by applying the data directly onto images in the form of rotation, flipping, and scaling so as to diversify the training data much further; hence, it generalizes well across several imaging conditions and potentially also patient demographics [35].

Transfer learning, on the training part, is relatively important. First, a Swin Transformer V2 model is pre-trained on a general image dataset with an aim to understand the basics of visual features. Pretraining allows models to gain strong baseline understanding of structures in images. Then, this model is fine-tuned with specific CXR data. In the process, pre-learned features would be fine-tuned to the specific nature of medical images, consequently improving the model's capacity to give very accurate diagnoses on pneumonia.

The model hyperparameters are optimized through fine-tuning using stochastic gradient descent techniques in reference to the betterment rate increase, batch size manageability, and training epochs reasonability. Last but not least, the model is cross-validated with data partitioning in order to check for unbiased performance against unseen cases.

The performance of Swin Transformer V2 is gauged on common metrics such as accuracy, precision, recall, and F1 score. High test accuracy ensures that most cases of pneumonia that the system diagnoses are indeed positive, with high precision and high recall making it specific and sensitive in its predictions. The F1 score is a way the harmonic mean of precision and recall can be visualized [37].

4.2 Ensemble Machine Learning Classifier for Financial Fraud Detection

Financial fraud in the health sector involves convoluted and changing challenges; therefore, it requires good, robust, and adaptive machine learning techniques. Ensemble methods in machine learning are ways of combining and putting into operation many classifiers in order to enhance performance in detection and, at the same time, reduce false positives.

The Ensemble Method joins various predictions given by many machine learning algorithms to achieve higher accuracy than any one of them alone. In this case, for this paper, an ensemble probably means the aggregated outputs of RF, NB, and DT into one set. This is extremely effective in the area of fraud detection, and in addition, the unique features such as feature importance and capability for balancing overfitting that RF offers, along with the probability-based predictions and NB, can be combined with the well-defined decision criteria in DTs.

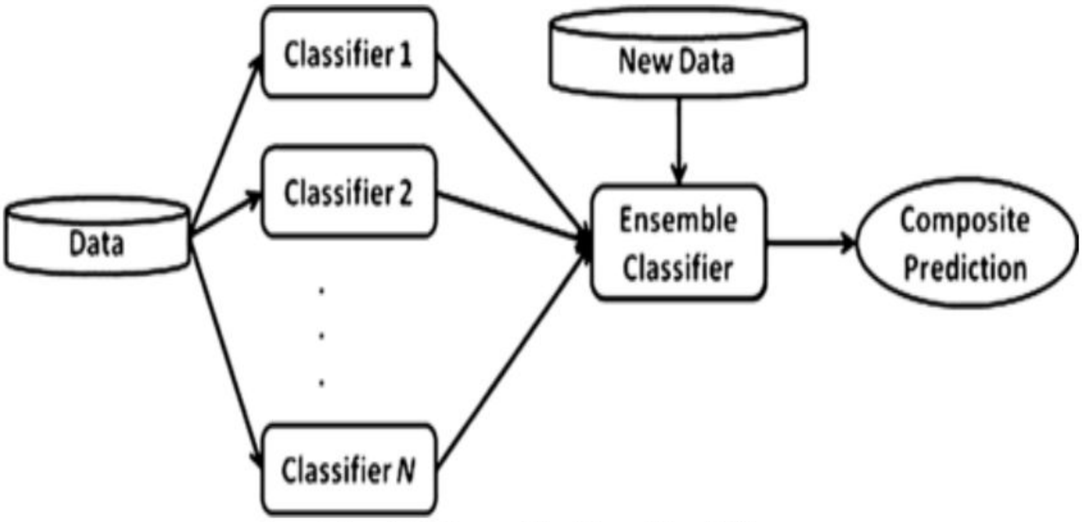


Fig. 4.2: Ensemble Classifier [38]

4.2.1 Random Forest (RF): Random Forest (RF) is a strong ensemble learning method applied extensively in classification and regression tasks. It grows many decision trees through the training set and outputs the class as a mode of the classes for the classification tasks, which are predicted by individual trees. Very strong since it reduces the risk of the overfitting, which is typical in single decision trees, because averaged several decision trees, trained on different fractions of one training set. Since this approach is based on averaging, it reduces variance and improves prediction. RF also allows for an interpretation of feature importance, which is useful for fraud detection in explaining important factors [39].

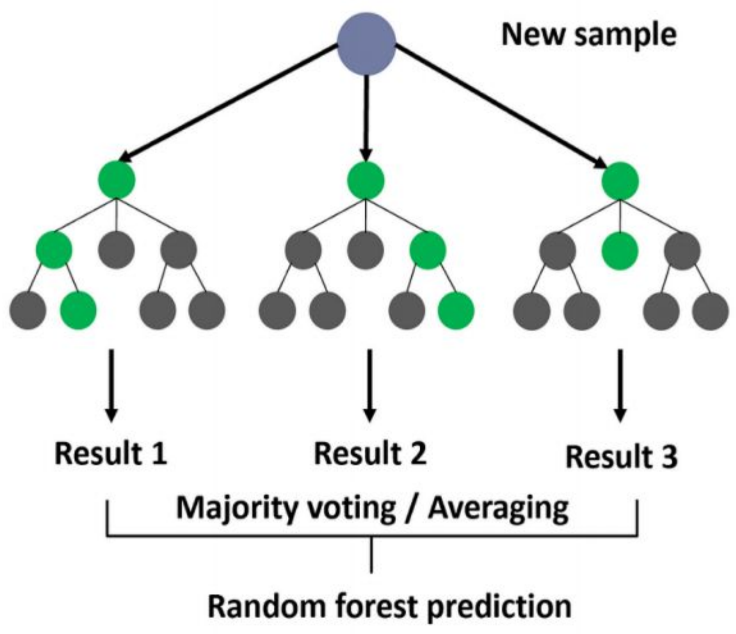


Fig. 4.3: Random Forest Working [40]

4.2.3 Naive Bayes (NB): Naive Bayes classifiers assume independence of predictors and calculate the probability that an event will occur given prior knowledge on the conditions that might be involved with the event. The classifiers are quite simple in nature, but they are very effective particularly with large datasets. It is very compatible with binary or multi-class classification tasks and finds prominent use in spam filtering, sentiment analysis, and even medical diagnostic tests, making it quite appropriate for pinpointing fraudulent transactions; after all, patterns can be based on historic data [41].

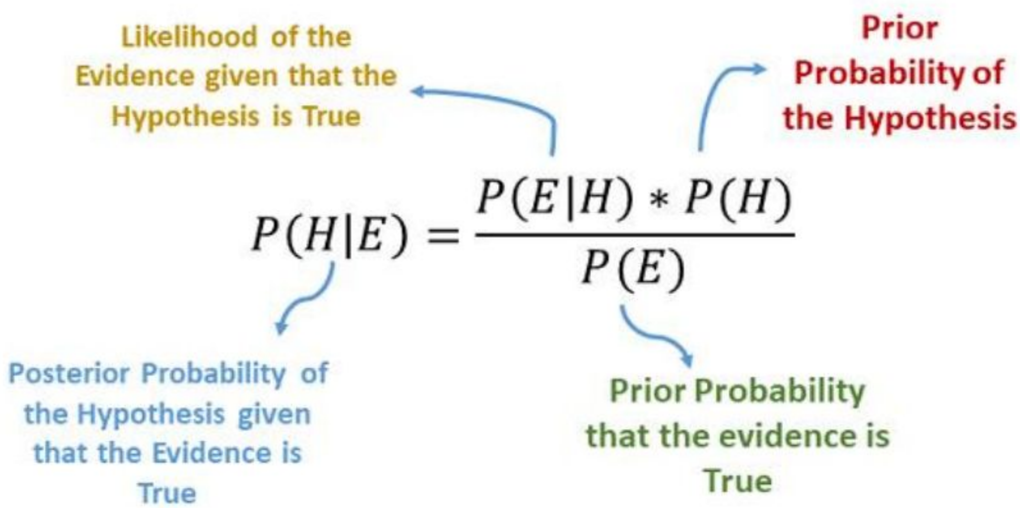


Fig. 4.4: Naïve Bayes [42]

4.2.4 K-Nearest Neighbors (KNN): The KNN classifier appears to be a simple and easy-to-use supervised machine-learning method which has the potential to be employed in solving of problems with regression and classification. Uses for this strategy include locating people who have made false claims on their auto insurance and locating those who have allowed their credit card payments to go behind [43]. The basic idea behind this method is that, given a certain number of K, this would identify the K closest classes to an unlabeled data point and assign that class to it based on which of those K classes has the most data points [43].

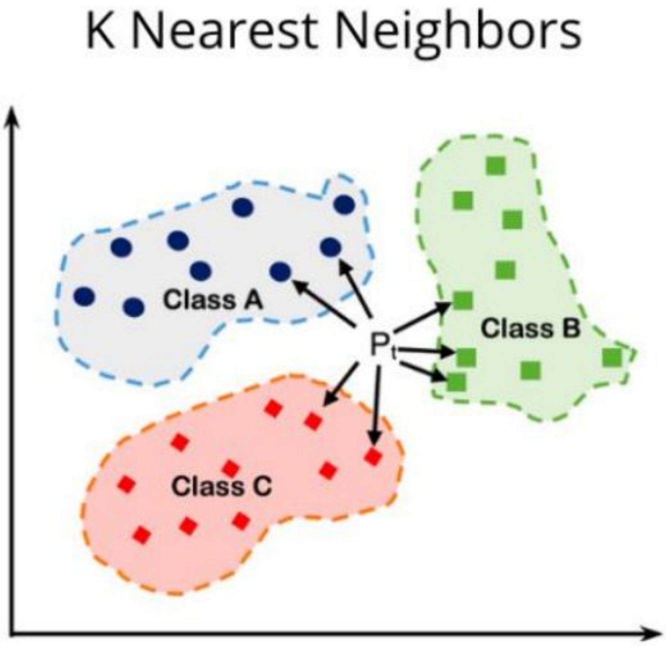


Fig. 4.5: KNN Architecture [44]

In conclusion, the techniques used in this study leverage advanced machine learning models to optimize healthcare diagnostics and protect financial integrity. The Swin Transformer V2 enhances diagnostic accuracy through its sophisticated architecture, while the ensemble machine learning classifier improves fraud detection by combining the strengths of multiple algorithms. Together, these techniques provide a comprehensive solution for improving patient care and safeguarding healthcare finances.

CHAPTER 5

RESULT and DISCUSSION

The conclusions that were derived from the results and discussions sections of this study clearly show the strong effects and positive contributions of using advanced machine learning in the field of healthcare diagnostic as well as the field of financial fraud prevention and detection. The application of such technologies as indicated in our discussion reflects a definite step in the right direction towards improving the efficiency of the healthcare system through improving diagnosing efficacy for life-threatening illnesses, specifically pneumonia and the ability to better prevent attempts at fraud in the healthcare payment system. This study then presents a comprehensive analysis of the impact of machine learning in addressing these two challenges that the healthcare and fraud detection industries face, through a thorough evaluation of a case using Swin Transformer V2 for medical imaging and Random Forest, Naïve Bayes, and Decision Trees for fraud detection. The following discourse goes ahead to further analyze the performance metrics of the models and their ability in achieving not only the advancement of clinical diagnosis that enhances better healthcare outcomes but also strengthening the financial side of the institutions hence contribute to the overall improvement of the health care system.

5.1 Financial Fraud Detection

In this study, we employed four ML classifiers: NB, RF, KNN, and Ensemble, to detect financial fraud within the healthcare domain. We used a dataset sourced from Kaggle to test and gauge the efficiency of these classifiers. The classifiers' performance was measured against criteria like Accuracy, Precision, and Recall. As depicted in Table 4, the ensemble classifier outperformed the others, achieving scores of 99.46, 98.38, and 98.58 in Accuracy, Precision, and Recall, respectively, surpassing the individual performances of NB, RF, and KNN.

TABLE 5.1: PERFORMANCE ANALYSIS FOR FOUR CLASSIFIERS.

Technique	**Accuracy**	**Precision**	**Recall**	**F1-Score**
NB	97.18	93.48	92.04	92.75
KNN	96.97	94.86	92.05	93.43
RF	98.66	97.58	97.78	97.67
Ensemble	99.46	98.38	98.58	98.47

In figure 5.1 the graphical representation of the overall performance of NB, RF, KNN, and Ensemble classifiers is illustrated.

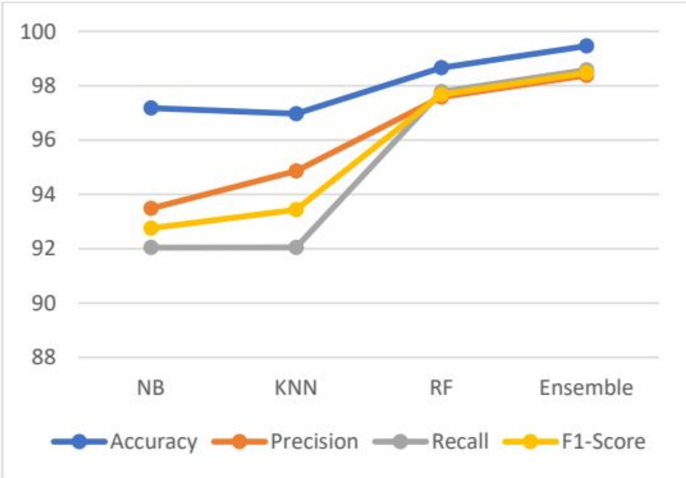


Fig. 5.1. Performance analysis

5.1.1 ROC Curve

Using threshold values from 0 to 1, Fig. 5.2 shows how the AUC (Area Under the Curve) is determined. In the provided graph, the blue lines which are the Ensemble technique represent a better scenario, indicating a model with higher TP rates and lower FP rates. The other two lines represent a mid-scenario that is RF and NB, indicating moderate performance. Finally, the red line which is KNN represents the worst scenario, indicating a model with low TP rates and high FP rates.

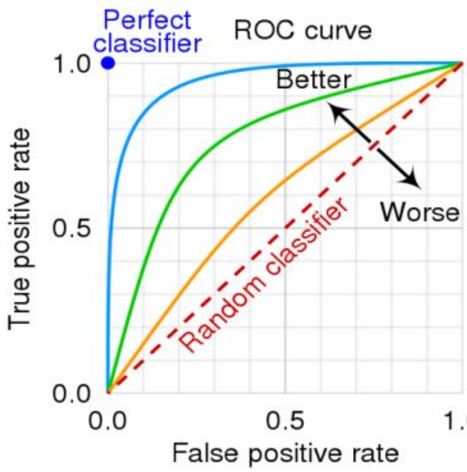


Fig. 5.2: ROC Curve for NB, KNN, RF, Ensemble

The choice of the ensemble method was pivotal to our study. This approach harnesses the power of multiple machines learning algorithms, thereby delivering enhanced performance through collective intelligence. However, we acknowledge limitations, including potential biases in data collection and the model's inclination toward more prevalent forms of fraud, possibly overlooking rare, anomalous fraudulent activities.

5.2. Pneumonia Diagnosis:

The pneumonia detection using the Swin Transformer V2 model on the CXR dataset yielded impressive results. The model achieved 98.6% accuracy, 98.5% precision, 98.1% recall, and a 98.3% F1-score, demonstrating its effectiveness. It outperformed other models, including AlexNet, MobileNetV3, and its predecessor, Swin Transformer V1. This improvement highlights the enhanced representational capabilities and spatial hierarchy learning of the Swin Transformer V2. AlexNet and VGG-16 showed accuracies of 97.2% and 95.3%, respectively, emphasizing their significance in image classification.

Table 5.2: Performance Analysis of Swin Across Different Models

Methods	Accuracy	Precision	Recall	F1-score
AlexNet	97.2	96.4	96	96.2
MobileNetV3	96	96.5	95.5	96
InceptionV3	89.5	90.2	89	89.6
VGG-16	95.3	95.5	95.1	95.3
Xception	83.5	84.3	83.2	83.7
EfficientNet B0	90.6	89.9	89.3	89.6
DenseNet	94.7	93.9	93.5	93.7
ResNet 50	95.1	95.6	95.1	95.3
Swin Transformer V1	96.4	95.9	95.4	95.6
Swin Transformer V2	98.6	98.5	98.1	98.3

While easier architectures like MobileNetV3 prioritize efficiency with a precision of 96.5%, deeper models inclusive of ResNet 50 and DenseNet gain accuracies over 94%, highlighting their effectiveness in characteristic extraction. However, the Swin Transformer V2 outperforms them in accomplishing a balanced precision-remember alternate-off, setting up itself as a benchmark for pneumonia detection. Its high sensitivity helps to minimize false positives, important in scientific diagnostics; its high specificity ensures that superb cases are not missed in truth.

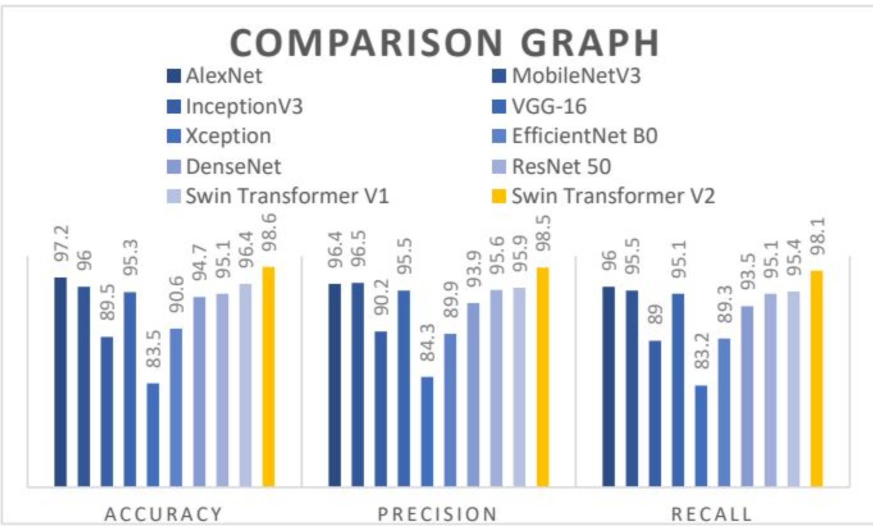


Fig. 5.3: Comparison of various models

The F1 score further cements the fact that the Swin Transformer V2 has a decent balance of intensity between two metrics: precision and recall. Such balance is rather important in medical image processing, since one needs to trade-off between trying to detect as many positive cases as possible with keeping false alert rates low.

Swin Transformer performs brilliantly based on the progressive structure: it adjusts window size and focusing mechanisms to capture V-tail microscopic styles of abnormalities in chest X-ray photographs, internally with indicators of pneumonia. The model processes photographs at different scales: super-consistent performance abnormalities are found with hyper-consistency. In summary, Swin Transformer V2 is a very good model for predicting pneumonia, better than the older models concerning over-accuracy and keep-in-mind rate. It is so promising to be used in scientific use because of the importance of accuracy.

The findings expose excellent optimization of tasks in the health care sector by the application of machine learning, particularly in improved diagnosis accuracy and protecting the integrity of financials. In medical diagnostics, the implementation of the Swin Transformer V2 model has seen vast improvements in the accurate diagnosis of pneumonias by means of chest X-ray images. A hierarchical structure of the model, reinforced with the use of shifted windows mechanism, was able to capture the complexity of low-level details and subtleties in high-resolution images for good diagnostic performance.

The training and validation of the Swin Transformer V2 has been done on a very wide and broad dataset pertaining to CXR images. The model has shown a very high rate of accuracy up to 98.6%, implying being able to predict most of the cases of pneumonia correctly. Further, the high precision of 98.5% indicates that the model is exceptionally good in false-positive minimization, along with a recall of 98.1% to ensure that none of the true positive cases is missed out on. The F1 score, the average of precision and recall, was 98.3%, stressing again that this model was robust and very useful in practical clinical applications. These results point to the potential of the previously shown model to increase the level of accuracy in diagnosis, which would ultimately result in better patient outcome through timely and accurate medical interventions.

The ensemble model combined Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbors. This was to allow each single classifier to maximize its strong points toward enhanced performance in fraud detection. Data from the dataset—train and validation—was used for building the ensemble model, which collected millions of financial transactions. For this, core attributes, like transaction amounts, origin and destination accounts, and transaction types, were detailed in a careful analysis to identify any patterns characteristic of fraud.

The ensemble classifier achieved an accuracy of 99.46% in the identification of legitimate and fraudulent transactions. The model's precision was at 98.38%, stating that the false-positive rate is low and the recall was 98.58%, meaning it could identify most of the activities as fraudulent. The high F1 score confirms properly that this model performs quite balanced detection of fraud in a way that possibly will ensure the model is highly specific and sensitive. At its core, the importance of such results is to protect the financing resources of the healthcare providers themselves, reduce the economic burden of fraud, and enhance overall financial security by reducing vulnerabilities within the healthcare domain.

Only two such prime examples are Swin Transformer V2, used for the diagnosis of medical problems, and the ensemble classifier, meant for detection of fraud. These show the power of transformation brought through the machinery of machine learning in health. The advanced models integrated by a provider present improvement facet that are critical both diagnostically and financially to health care. These research findings have created a stimulus about the effectiveness of machine learning techniques in solving challenges that are considered cumbersome and aimed at contributing toward the optimization of delivering health care.

Taken together, this study presents the tremendous benefits of applying machine learning in healthcare. While Swin Transformer V2 has been directed to greater diagnostic accuracy, the use of the ensemble classifier repeatedly attests to the guarding of financial integrity. These results underscore the need for improved development and deployment of advanced machine-learning models to enhance clinical and financial outcomes in health care.

Provided below is the table summarizing the results of the present study, showing the performance of Swin Transformer V2 and ensemble classifier for financial fraud detection:

TABLE 5.3: SUMMARIZING THE RESULT

Model	**Task**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Swin Transformer V2	Pneumonia Diagnosis (CXR Images)	98.6	98.5	98.1	98.3
Ensemble Classifier	Financial Fraud Detection	99.46	98.38	98.58	98.47

This table captures the key performance metrics, highlighting the high accuracy, precision, recall, and F1 scores achieved by both models in their respective tasks. The results demonstrate the effectiveness of the Swin Transformer V2 in accurately diagnosing pneumonia from chest X-ray images and the ensemble classifier in detecting financial fraud within healthcare transactions.

CHAPTER 6

CONCLUSION, FUTURE SCOPE and SOCIAL IMPACT

All in all, the thesis offered herein has very well earned the utility of machine learning techniques to optimize health through the enhancement of diagnostic accuracies and to improve financial fraud detection mechanisms. This twin approach has provided broad insights on the potential for machine learning to change healthcare operations by enhancing not only methodologies of medical diagnoses but also by adding towards the building of stronger financial systems within the health facilities against fraudulent activities.

This study proves that the Swin Transformer V2 holds promise for the diagnosis of pneumonia from chest X-ray snapshots. It is with these advanced architectural designs, including hierarchical systems and shifted window mechanisms, where it has shown this model can really be sensitive to subtle and complex signatures existing in clinical imaging with high accuracy. The results from the diagnostic phase of this study suggest that the ability of the model to produce truly impressive diagnostic accuracies is to play a significant role in clinical practice, with a view to improving patient outcomes by allowing timely and accurate diagnoses.

Parallel to the advancements in scientific diagnostics, this thesis additionally explored the efficacy of an ensemble technique in detecting monetary fraud inside healthcare transactions. By integrating several sturdy algorithms—Random Forest, Naive Bayes, and Decision Trees—the ensemble model efficiently navigated the elaborate styles of healthcare transactions to become aware of fraudulent activities efficaciously. This part of the examine underscored the critical position that gadget gaining knowledge of can play in safeguarding the monetary elements of healthcare, therefore preventing sizeable monetary losses through early detection of fraud.

The integration of these device mastering technologies has the capability to significantly lessen the burden on healthcare specialists and directors through imparting them with reliable, efficient, and automated gear for medical prognosis and fraud detection. The convergence of these technology with present healthcare practices can lead to more targeted affected person care practices, optimized operational efficiencies, and more desirable security features, culminating in a better healthcare gadget.

In end, this thesis has verified that machine studying is an invaluable asset in the pursuit of optimized healthcare delivery. The technology investigated right here—in particular the Swin Transformer V2 for medical imaging and the ensemble device getting to know fashions for fraud detection—gift compelling proof that system studying may be successfully integrated into diverse facets of healthcare. Future work has to attention on expanding this technology to different regions of healthcare, exploring extra illnesses and broader economic applications, and continually refining the fashions to adapt to the evolving landscape of healthcare challenges. The route forward involves not only technological development however also a deep integration of these equipment into normal healthcare strategies to recognize the overall capability of system getting to know in optimizing healthcare effects and operational efficiencies.

As this thesis advances our understanding of the intersection between machine mastering and healthcare optimization, future research must consider numerous avenues to beautify and extend upon the modern findings. There is significant capability to increase the utility of the Swin Transformer V2 version and the ensemble gadget studying technique to other diagnostic regions beyond pneumonia, including cardiology or oncology, wherein imaging plays a crucial position in prognosis. Further, the version of those models to combine actual-time records processing can provide immediately insights at some point of clinical assessments, potentially revolutionizing patient care and reaction instances.

Additionally, exploring deeper integration of gadget gaining knowledge of models with digital health facts (EHR) should provide more comprehensive insights into affected person history, enhancing diagnostic accuracy and personalized treatment plans. The capability to investigate longitudinal affected person statistics thru superior algorithms might are expecting patient outcomes extra efficaciously and advocate preventative measures. Moreover, expanding the dataset to consist of multi-institutional records should decorate the robustness and generalizability of the models, ensuring they perform nicely throughout numerous healthcare settings and populations.

The societal impact of applying advanced machine learning in healthcare: through the increased accuracy of diagnostics, machine learning technology can considerably contribute to reducing misdiagnoses and, therefore, securing timely treatment of patients with the most appropriate methods, eventually saving lives and lowering costs for healthcare. On the monetary front, robust models for fraud detection secure economic resources and assure stakeholders that budgets are spent on offering patient care and not just made to disappear through fraud.

Another important issue regarding the use and application of systems learning solutions in health care has to do with ethical issues and policy implications. This is vital in establishing trust and acceptability: the privacy of data, the security of information regarding patients, and the transparency of the methodology. The development of recommendations and guidelines that govern the ethical use of AI in health will have to be realized as these technologies increasingly underpin the medical and administrative processes.

In conclusion, the growth of system studying packages in healthcare holds splendid promise no longer only for enhancing medical diagnostics and financial security but additionally for its capacity to supply more equitable and efficient healthcare services globally. The integration of these technology ought to result in a paradigm shift in how affected person care is delivered and managed, fostering a more sustainable and affected person-focused healthcare machine. Future studies and improvement will play a important position in knowing the full capability of machine gaining knowledge of to convert healthcare in both scientific and social dimensions.

REFERENCES

[1] Ourworldindata.org. [cited 2024 Feb 23]. Available from: https://ourworldindata.org/pneumonia.

[2] Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: general overview. Korean journal of radiology, 18(4), 570-584.

[3] Gabruseva, T., Poplavskiy, D., & Kalinin, A. (2020). Deep learning for automatic pneumonia detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 350-351).

[4] Sadgali, Imane, Nawal Sael, and Faouzia Benabbou. "Performance of machine learning techniques in the detection of financial frauds." Procedia computer science 148 (2019): 45-54.

[5] Amiram, Dan, Zahn Bozanic, James D. Cox, Quentin Dupont, Jonathan M. Karpoff, and Richard Sloan. "Financial reporting fraud and other forms of misconduct: a multidisciplinary review of the literature." Review of Accounting Studies 23 (2018): 732-783.

[6] Modi, Krishna, and Reshma Dayma. "Review on fraud detection methods in credit card transactions." In 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1-5. IEEE, 2017.

[7] Zhai, X., Kolesnikov, A., Houlsby, N., & Beyer, L. (2022). Scaling vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12104-12113).

[8] Amin, R., Al Ghamdi, M. A., Almotiri, S. H., & Alruily, M. (2021). Healthcare techniques through deep learning: issues, challenges and opportunities. IEEE Access, 9, 98523-98541.

[9] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[10] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[13] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

[14] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. cell, 172(5), 1122-1131.

[15] Elfadil Eisa, Arafat Al-Dhaqm, Maged Nasser, Tusneem Elhassan, Hashim Elshafie, and Abdu Saif. "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review." Applied Sciences 12, no. 19 (2022): 9637.

[16] Amponsah, Anokye Acheampong, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology." Decision Analytics Journal 4 (2022): 100122.

[17] Yıldırım Taşer, Pelin, and Fatma Bozyiğit. "Machine Learning Applications for Fraud Detection in Finance Sector." In The Impact of Artificial Intelligence on Governance, Economics and Finance, Volume 2, pp. 121-146. Singapore: Springer Nature Singapore, 2022.

[18] Turaba, Mahbuba Yesmin, Mehedi Hasan, Nazrul Islam Khan, and Hafiz Abdur Rahman. "Fraud Detection During Financial Transactions Using Machine Learning and Deep Learning Techniques." In 2022 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), pp. 1-8. IEEE, 2022.

[19] Mehbodniya, Abolfazl, Izhar Alam, Sagar Pande, Rahul Neware, Kantilal Pitambar Rane, Mohammad Shabaz, and Mangena Venu Madhavan. "Financial fraud detection in healthcare using machine learning and deep learning techniques." Security and Communication Networks 2021 (2021): 1-8.

[20] Asha, R. B., and Suresh Kumar KR. "Credit card fraud detection using artificial neural network." Global Transitions Proceedings 2, no. 1 (2021): 35-41.

[21] Azhan, Mohammed, and Shazli Meraj. "Credit card fraud detection using machine learning and deep learning techniques." In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 514-518. IEEE, 2020.

[22] Masood, Anum, Usman Naseem, and Imran Razzak. "Multi-scale swin transformer enabled automatic detection and segmentation of lung metastases using CT images." 2023 IEEE 20th international symposium on biomedical imaging (ISBI). IEEE, 2023.

[23] Wu, Xinli, et al. "Multi-Scale Swin Transformer for COVID-19 Classification in Chest X-Rays." 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2023.

[24] Ramli, Anis Amirah Binti, et al. "Automatic Pneumonia Detection Through Chest X-Ray Image-Based." 2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS). IEEE, 2023.

[25] Ye, Zixun, Hongying Zhang, and Qiuyang Liu. "SWTFace: A Multi-branch Network for Masked Face Detection and Recognition." 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI). IEEE, 2022.

[26] Yaseliani, Mohammad, et al. "Pneumonia detection proposing a hybrid deep convolutional neural network based on two parallel visual geometry group architectures and machine learning classifiers." IEEE Access 10 (2022): 62110-62128.

[27] Li, Fangfang, et al. "Intelligent Pneumonia Detection Model of Focal Transformer and Multi-scale-input SPD." 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2022.

[28] Nahiduzzaman, Md, et al. "A novel method for multivariant pneumonia classification based on hybrid CNN-PCA based feature extraction using extreme learning machine with CXR images." IEEE Access 9 (2021): 147512-147526.

[29] Zhang, Jianpeng, et al. "Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection." IEEE transactions on medical imaging 40.3 (2020): 879-890.

[30] Sharma, Harsh, et al. "Feature extraction and classification of chest x-ray images using cnn to detect pneumonia." 2020 10th international conference on cloud computing, data science & engineering (Confluence). IEEE, 2020.

[31] Ferreira, José Raniery, et al. "Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest x-ray images." 2020 42nd annual international conference of the IEEE engineering in Medicine & Biology Society (EMBC). IEEE, 2020.

[32] Labhane, Gaurav, et al. "Detection of pediatric pneumonia from chest X-ray images using CNN and transfer learning." 2020 3rd international conference on emerging technologies in computer engineering: machine learning and internet of things (ICETCE). IEEE, 2020.

[33] Alcázar, C. (no date) Alkzar90/NIH-chest-X-ray-dataset · datasets at hugging face, alkzar90/NIH-Chest-X-ray-dataset · Datasets at Hugging Face. Available at: https://huggingface.co/datasets/alkzar90/NIH-Chest-X-ray-dataset (Accessed: 21 May 2024).

[34] Dr. Liji Thomas, M. (2020) Transfer learning exploits chest-xray to diagnose COVID-19 pneumonia, News. Available at: https://www.news-medical.net/news/20201218/Transfer-learning-exploits-chest-Xray-to-diagnose-COVID-19-pneumonia.aspx (Accessed: 21 May 2024).

[35] Jian, S., Kaiming, H., Shaoqing, R., & Xiangyu, Z. (2016). Deep residual learning for image recognition. In IEEE Conference on Computer Vision & Pattern Recognition (pp. 770-778).

[36] Z. Liu et al., “Swin Transformer V2: Scaling up capacity and Resolution,” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022. doi:10.1109/cvpr52688.2022.01170

[37] Devnath, L., Luo, S., Summons, P., & Wang, D. (2021). Automated detection of pneumoconiosis with multilevel deep features learned from chest X-Ray radiographs. Computers in Biology and Medicine, 129, 104125.

[38] Bramer, Max, and Max Bramer. "Ensemble classification." Principles of Data Mining (2013): 209-220.

[39] Heuvelink, and Benedikt Gräler. "Random forest as a generic framework for predictive modeling of spatial and Spatiotemporal variables." PeerJ 6 (2018): e5518

[40] Yehoshua, Dr.R. (2023) Random forests, Medium. Available at: https://medium.com/@roiyeho/random-forests-98892261dc49 (Accessed: 21 May 2024).

[41] Wei, J., Yi, L., Giacoumidis, E., Cheng, Q., & Tao Lau, A. P. (2020). Special Issue on “Optics for AI and AI for Optics”. Applied Sciences, 10(9), 3262

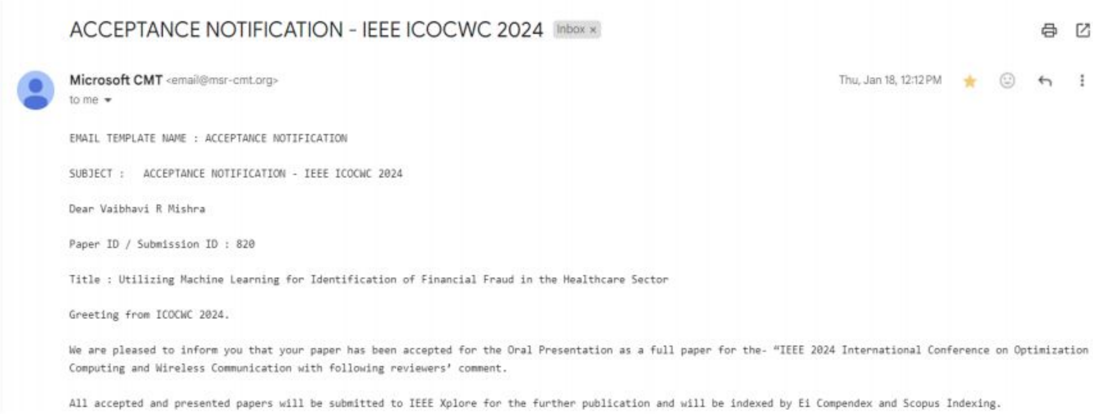
[42] Gamal, B. (2024) Naïve Bayes algorithm, Medium. Available at: https://medium.com/analytics-vidhya/na%C3%AFve-bayes-algorithm-5bf31e9032a2 (Accessed: 21 May 2024).

[43] I-King-Of-Ml (2019) KNN(K-nearest neighbour) algorithm, maths behind it and how to find the best value for K, Medium. Available at: https://medium.com/@rdhawan201455/knn-k-nearest-neighbour-algorithm-maths-behind-it-and-how-to-find-the-best-value-for-k-6ff5b0955e3d (Accessed: 21 May 2024).

[44] Sachinsoni (2023) K nearest neighbours-introduction to machine learning algorithms, Medium. Available at: https://medium.com/@sachinsoni600517/k-nearest-neighbours-introduction-to-machine-learning-algorithms-9dbc9d9fb3b2 (Accessed: 21 May 2024).

LIST OF PUBLICATIONS

1. *Ruchika Malhotra, Vaibhavi Rajesh Mishra, "Utilizing Machine Learning for Identification of Financial Fraud in the Healthcare Sector"*. The paper has been **Accepted** as well as **Published** at the 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC). Indexed by **Scopus**. Paper Id: 820



ACCEPTANCE NOTIFICATION - IEEE ICOCWC 2024 Inbox ×

Microsoft CMT <email@msr-cmt.org>
to me

Thu, Jan 18, 12:12 PM

EMAIL TEMPLATE NAME : ACCEPTANCE NOTIFICATION

SUBJECT : ACCEPTANCE NOTIFICATION - IEEE ICOCWC 2024

Dear Vaibhavi R Mishra

Paper ID / Submission ID : 820

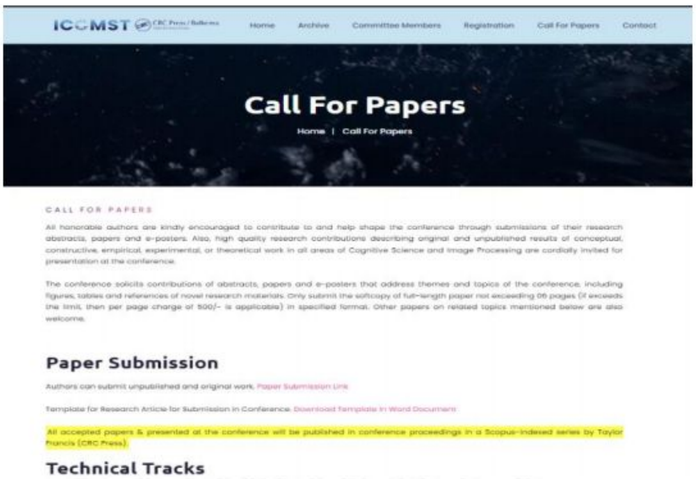
Title : Utilizing Machine Learning for Identification of Financial Fraud in the Healthcare Sector

Greeting from ICOCWC 2024.

We are pleased to inform you that your paper has been accepted for the Oral Presentation as a full paper for the- "IEEE 2024 International Conference on Optimization Computing and Wireless Communication with following reviewers' comment.

All accepted and presented papers will be submitted to IEEE Xplore for the further publication and will be indexed by Ei Compendex and Scopus Indexing.

2. *Vaibhavi Rajesh Mishra, Ruchika Malhotra, "Empowering Healthcare with Swin Transformer V2: Advancing Pneumonia Diagnosis through Deep Learning"*. The paper has been **Accepted** at 4th International Conference on Computational Methods in Science and Technology 2024 (ICCMST). Indexed by **Scopus**. Paper Id: 52



(Paper id : 52) Acceptance of paper for 4th International Conference on Computational Methods in Science and Technology ICCMST 2024 Inbox ×

Microsoft CMT <email@msr-cmt.org> Thu, Mar 21, 4:38 PM

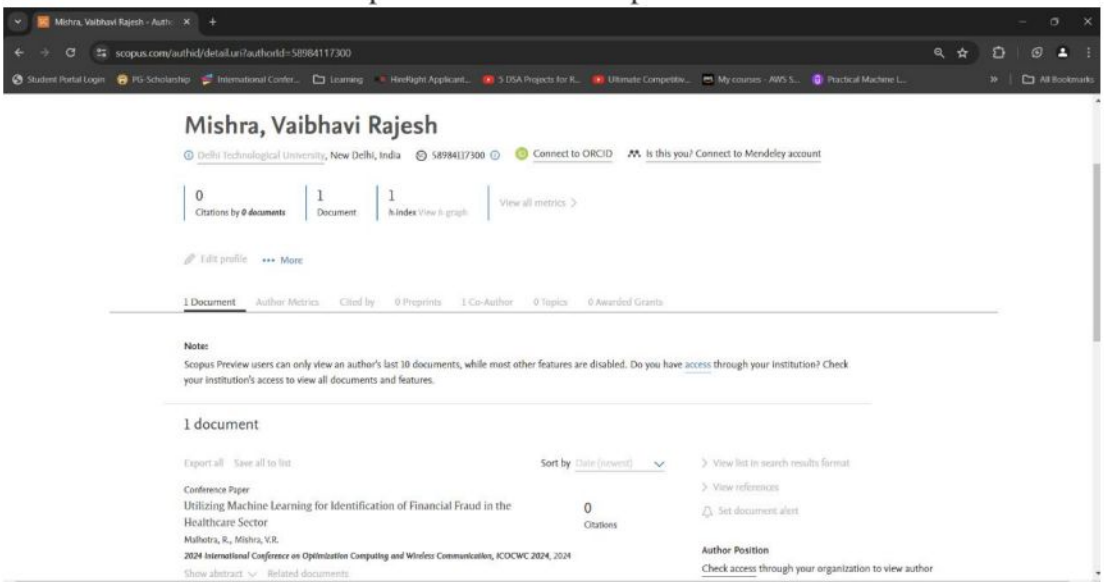
to me

Dear Dr./Mr./Ms. Vaibhavi R Mishra,

Congratulations!!!

As per reviewer's comment, on behalf of the ICCMST 2024 program committee and technical committee, we are very pleased to inform you that Submission id : 52 with Title:- Empowering Healthcare with Swin Transformer V2: Advancing Pneumonia Diagnosis through Deep Learning has been accepted with minor revisions as a REGULAR paper for presentation at the 4th International Conference on Computational Methods in Science and Technology dated 2nd - 3rd May 2024.

3. Paper 1 titled "*Utilizing Machine Learning for Identification of Financial Fraud in the Healthcare Sector*" is published under Scopus indexed.



scopus.com/authid/detail.uri?authorId=58984117300

Mishra, Vaibhavi Rajesh

Delhi Technological University, New Delhi, India 58984117300 Connect to ORCID Is this you? Connect to Mendeley account

0 Citations by 0 documents | 1 Document | 1 h-index View h-graph | View all metrics

Edit profile ••• More

1 Document | Author Metrics | Cited by | 0 Preprints | 1 Co-Author | 0 Topics | 0 Awarded Grants

Note:

Scopus Preview users can only view an author's last 10 documents, while most other features are disabled. Do you have access through your institution? Check your institution's access to view all documents and features.

1 document

Export all Save all to list Sort by Date (newest)

Conference Paper

Utilizing Machine Learning for Identification of Financial Fraud in the Healthcare Sector

0 Citations

Malhotra, R., Mishra, V.R.

2024 International Conference on Optimization Computing and Wireless Communication, ICOCWC 2024, 2024

Show abstract Related documents

View list in search results format

View references

Set document alert

Author Position

Check access through your organization to view author