Osteoporosis Disease Detection Using Transfer Learning

Project Report Submitted to National Institute of Technology, Jamshedpur in Partial Fulfilment of the Requirements for the Award of Degree of

Master of Computer Applications (2021-2024)

Submitted by

Shreyansh Shukla (2021PGCACA035)

Under the esteemed guidance of

Dr. Jitesh Pradhan
Assistant Professor

Department of Computer Science and Engineering
National Institute of Technology, Jamshedpur – 831014, Jharkhand, INDIA

राष्ट्रीय प्रौद्योगिकी संस्थान जमशेदपुर
जमशेदपुर – 831014, झारखण्ड, भारत
(राष्ट्रीय महत्व की संस्था, शिक्षा मंत्रालय, भारत सरकार)
NATIONAL INSTITUTE OF TECHNOLOGY JAMSHEDPUR
Jamshedpur – 831014, Jharkhand, India
(An Institution of National Importance under Ministry of Education, Government of India)

CERTIFICATE

This is to certify that the report entitled **Osteoporosis Disease Detection Using Transfer Learning** submitted by **Shreyansh Shukla (2021PGCACA035)** under the guidance of **Dr. Jitesh Pradhan,** Assistant Professor of Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jharkhand in partial fulfilment of the requirement for the award of degree of **Master of Computer Applications (MCA)** during to academic session 2021-24 to **National Institute of Technology, Jamshedpur.**

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein but approve the project report for the purpose for which it is submitted.

DR. DANISH ALI KHAN

(Professor)

Head of the Department

Dept. of Computer Science and Engineering

National Institute of Technology, Jamshedpur

DR. JITESH PRADHAN

(Assistant Professor)

Supervisor

Dept. of Computer Science and Engineering

National Institute of Technology, Jamshedpur

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referred the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misinterpreted or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be the cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

Shreyansh Shukla

(2021PGCACA035)

Master of Computer Applications
2021-24

National Institute of Technology,
Jamshedpur

ACKNOWLEDGEMENT

In this infinite universe of mysteries beyond known and unknown, the teacher is one who shows the path to explore by illuminating with the knowledge he has, thus dispelling out the darkness and fear in one's mind. I, therefore, owe my sincere thanks to my guide, **Dr. Jitesh Pradhan,** Assistant Professor**,** Department of Computer Science and Engineering, National Institute of Technology, Jamshedpur for kindling in me the love for this beautiful subject. His constant suggestions and feedbacks enabled me to bring out fruitful results, thus exploring new areas and unveiling new possibilities. I am truly grateful for their constant encouragement.

A big thank you to all the faculty members of the department for their kind assistance and cooperation, which played a vital role in shaping the project.
Lastly, heartfelt thanks to my friends for their support and contributions in making this project a success.

I am grateful to all my teachers, from KG to PG, for all the subjects they have taught me, thus, making me a complete human-being.

This project would not have been possible without the collective efforts of these individuals, and I am truly thankful for their contributions.

ABSTRACT

Osteoporosis is a chronic condition characterized by weakened bones, making individuals more susceptible to fractures, especially as they age. Detecting osteoporosis early is crucial for effective treatment and prevention of debilitating fractures. In this project, we focus on leveraging advanced machine learning techniques to enhance the detection of osteoporosis using medical imaging data.

We employ transfer learning, a powerful technique in deep learning, which involves leveraging knowledge from pre-trained models on large datasets to improve performance on a specific task—in this case, detecting osteoporosis from medical bone images. We utilize two state-of-the-art deep learning architectures: EfficientNetB2vo and VGG16, known for their effectiveness in image recognition tasks.

Our approach involves extracting informative features from medical bone images using the pre-trained models. These features capture important patterns and characteristics in the images that are indicative of osteoporosis. By leveraging transfer learning, we can benefit from the wealth of knowledge already embedded in these models, thus enhancing the accuracy and efficiency of our detection system.

To further enhance the diagnostic capabilities, we employ a variety of classification algorithms, including Decision Trees, Random Forests, Support Vector Machines, Naive Bayes, Logistic Regression, and AdaBoost. Each algorithm offers unique advantages and trade-offs, allowing us to explore different avenues for osteoporosis detection.

The efficacy of each model-algorithm combination is rigorously evaluated to identify the most accurate diagnostic model for early osteoporosis detection. By comparing the performance of various combinations, we aim to uncover insights into the most effective approach for identifying individuals at risk of osteoporosis, ultimately contributing to improved patient outcomes and healthcare practices.

CONTENTS

i. TABLE OF FIGURES

ii. TABLE OF TABLES

1. INTRODUCTION

1.1. BACKGROUND –

Overview of osteoporosis and the importance of early detection.

1.1.1. Introduction to Osteoporosis:

Osteoporosis is a debilitating skeletal condition characterized by reduced bone density and strength, rendering bones brittle and susceptible to fractures. Often referred to as a "silent disease," osteoporosis progresses silently, with no apparent symptoms until a fracture occurs. Comparable to a building losing its structural integrity, bones affected by osteoporosis become fragile, increasing the risk of fractures even from minor falls or bumps. This condition poses a significant public health concern globally, affecting millions of individuals and imposing substantial burdens on healthcare systems and economies.

Beyond its sheer prevalence, the impact of osteoporosis extends far beyond physical consequences. Fractures resulting from osteoporosis can lead to chronic pain, disability, loss of independence, and even premature mortality. Moreover, the economic burden associated with osteoporotic fractures is substantial, encompassing healthcare costs, lost productivity, and diminished quality of life.

Despite its pervasive nature, osteoporosis diagnosis presents challenges, primarily due to its asymptomatic nature in the early stages. Traditional diagnostic methods, such as bone density scans, have limitations in detecting subtle changes indicative of osteoporosis. Moreover, access to diagnostic tools and resources may be limited, particularly in underserved communities.

Early detection of osteoporosis is crucial for instituting preventive measures and mitigating fracture risk. Timely identification of individuals at risk enables targeted interventions aimed at preserving bone health and reducing the incidence of fractures. Medical imaging plays a pivotal role in osteoporosis diagnosis, providing clinicians with valuable insights into bone density and structure. Recent advancements in machine learning offer promising opportunities to enhance osteoporosis detection and management, paving the way for more efficient and accurate diagnostic approaches.

1.1.2. Prevalence and Impact:

Osteoporosis affects millions of individuals worldwide, posing a significant public health concern due to its associated morbidity and mortality. Fractures resulting from weakened bones can cause debilitating pain, impair mobility, and compromise overall quality of life. Moreover, osteoporotic fractures are associated with heightened mortality risk, underscoring the severity of the condition and its far-reaching consequences.

1.1.3. Challenges in Diagnosis:

Diagnosing osteoporosis presents inherent challenges, primarily due to its asymptomatic nature and the limitations of conventional diagnostic methods. While specialized tests like X-rays and bone density scans are available for diagnosis, they may not always be readily accessible or affordable for all individuals, hindering timely detection and intervention.

1.1.4. Importance of Early Detection:

Early detection of osteoporosis is paramount as it enables proactive measures to prevent fractures and preserve bone health. Similar to addressing structural weaknesses in a building before catastrophic failure, identifying osteoporosis in its early stages allows for the implementation of preventive strategies, such as lifestyle modifications and pharmacological interventions, to mitigate fracture risk and maintain bone strength.

1.1.5. Role of Medical Imaging:

Medical imaging plays a pivotal role in the diagnosis and monitoring of osteoporosis. By providing visual insights into bone density, structure, and integrity, imaging modalities like X-rays, dual-energy X-ray absorptiometry (DXA), and magnetic resonance imaging (MRI) aid clinicians in assessing bone health and identifying signs of osteoporosis.

1.1.6. Advancements in Machine Learning:

Advancements in machine learning have revolutionized medical imaging analysis, offering innovative solutions for more accurate and efficient disease detection. By leveraging computational algorithms, researchers can train computer systems to recognize patterns and abnormalities in medical images, facilitating earlier and more precise diagnosis of conditions like osteoporosis.

1.1.7. Rationale for the Study:

Against this backdrop, the present study seeks to harness the power of machine learning algorithms to enhance osteoporosis detection. By leveraging machine learning techniques to analyse medical images, the aim is to expedite the diagnostic process, improve detection accuracy, and ultimately enable timely intervention to mitigate the impact of osteoporosis-related fractures. Through interdisciplinary collaboration between computer scientists, clinicians, and imaging experts, this research endeavours to contribute to advancements in osteoporosis management and improve patient outcomes.

So, after knowing this we can understand why osteoporosis detection is important and how machine learning can make a difference.

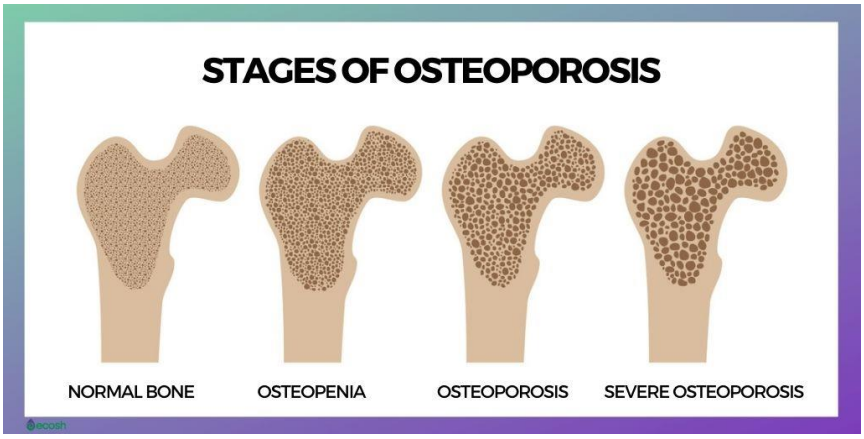

Fig 1. Stages of Osteoporosis

1.2. PROBLEM STATEMENT –

Osteoporosis is a prevalent skeletal disorder characterized by reduced bone density and increased vulnerability to fractures, particularly among aging populations. Timely detection of osteoporosis is essential for implementing preventive measures and reducing the risk of debilitating fractures. However, conventional diagnostic approaches, such as dual-energy X-ray absorptiometry (DEXA), can be expensive, time-consuming, and may not always yield accurate results. Additionally, these methods often require specialized equipment and expertise, limiting their accessibility, particularly in resource-constrained settings.

The lack of efficient and widely accessible diagnostic tools for osteoporosis poses a significant challenge in identifying individuals at risk and initiating appropriate interventions promptly. Furthermore, existing approaches may not fully leverage advancements in technology and machine learning, which could potentially enhance the accuracy and efficiency of osteoporosis detection.

Therefore, the problem at hand is twofold: firstly, **the need for more accessible and cost-effective methods for diagnosing osteoporosis**, and secondly, **the opportunity to leverage advancements in technology**, **particularly machine learning algorithms, to improve the accuracy and efficiency of osteoporosis detection**. Addressing these challenges could lead to earlier detection, better management, and ultimately, improved outcomes for individuals affected by osteoporosis.

1.3. OBJECTIVE –

Our primary objective is to develop a robust and accurate diagnostic model for osteoporosis detection, leveraging transfer learning techniques with deep learning architectures, we will focus and select to **EfficientNetV2B0** and **VGG16** models. By harnessing the pre-trained representations learned by these models on large-scale image datasets, we aim to extract

informative features from medical bone images that can effectively discriminate between healthy and osteoporotic bone tissues.

Additionally, we seek to explore the performance of various classification algorithms, including Decision Trees, Random Forests, Support Vector Machines, I Bayes, Logistic Regression, and AdaBoost, when combined with the pre-trained deep learning models. Through rigorous evaluation, we intend to assess the efficacy of each model-algorithm combination in terms of accuracy, sensitivity, specificity, and the area under the ROC curve (AUC), with the overarching goal of identifying the most effective approach for early osteoporosis detection.

Furthermore, we aim to compare the performance of different models and algorithms to determine the optimal diagnostic strategy for osteoporosis detection. Factors such as computational efficiency, interpretability, and scalability will be considered in this comparative analysis to ensure the practical feasibility and utility of the proposed diagnostic framework.

In addition to model development and evaluation, we plan to validate the selected diagnostic model on independent datasets to assess its generalizability and reliability across diverse patient populations and imaging modalities. This validation step is crucial for establishing the real-world applicability of the developed diagnostic tool and ensuring its robust performance across varied clinical settings. Finally, we endeavour to provide insights into the potential clinical impact of the developed diagnostic model, elucidating its implications for early intervention, fracture prevention, and overall patient care in the context of osteoporosis management. By highlighting the clinical relevance and utility of our diagnostic framework, we aim to contribute to improved patient outcomes and enhanced healthcare practices in the field of osteoporosis diagnosis and management.

2. LITERATURE REVIEW

2.1. OSTEOPOROSIS DETECTION USING MACHINE LEARNING TECHNIQUES AND FEATURE SELECTION [1]

Osteoporosis is the prevailing bone's disease, and its features are low bone density mass and the modification of their micro-architecture structure, so that bones' tolerance is reduced and the risk of fracture is increased. In osteoporosis, the Bone Mineral Density (BMD) is reduced; the bone micro-architecture is disrupted whereas the concentration and the variety of proteins in bones are altered. The classic osteoporotic fractures are hip, vertebral and wrist fractures. Osteoporotic fractures are defined as occurring at an associated with low BMD and which at the same time increased in incidence after the age of 50 years.

Apart from the direct physical implications of a fracture, such as pain and inconvenience, osteoporotic fractures are a major cause of morbidity and mortality. The lifetime risk in the United States for a hip, spine, or forearm fracture at the age of 50 years has been estimated to be 40% in women and 13% in men. In Sweden, the corresponding percentages are 46% for women and 22% for men. Caucasians and Asians are at increased risk, since African and Americans have 6% higher BMD. In the European Union one person breaks a bone because of osteoporosis every fifteen seconds. It is a fact that a percentage as high as 75% of the women with osteoporosis disregards this disorder. There are two types of osteoporosis, the primary (idiopathic) osteoporosis, which is a most frequent disease for women after menopause and is called postmenopausal osteoporosis. This type also includes the senile osteoporosis that may also be developed in men. The secondary osteoporosis, which may occur on anyone in the presence of particular hormonal disorders and other chronic diseases, as a result of medications, specifically glucocorticoids or other conditions causing increased bone loss by various mechanisms. In this case the disease is called steroid or glucocorticoid induced osteoporosis.

Often the first apparent symptom of osteoporosis is a broken bone, which is why the condition is also known as **"The Silent Crippler"**, as people do not realize they have osteoporosis until it's too late. However early detection and treatment of osteoporosis can decrease the fracture risk of a person to a minimum. For these reasons, there are studies using Artificial Intelligence techniques that are used for predicting whether a person has osteoporosis or not.

2.2. OPPORTUNISTIC OSTEOPOROSIS SCREENING USING CHEST RADIOGRAPHS WITH DEEP LEARNING: DEVELOPMENT AND EXTERNAL VALIDATION WITH A COHORT DATASET [2]

Osteoporosis and osteoporotic fractures have become global health issues of major concern owing to their association with age-related fractures in the aging countries including South Korea. By 2020, approximately 12.3 million people aged ≥50 years in the United States are expected to develop osteoporosis. Prevalence of osteoporosis in South Korea is 7.3% in males

and 38.0% in females aged ≥50 years by the data Fourth Korea National Health and Nutrition Examination Survey (KNHANES IV) 2008–2011. Osteoporosis is a systemic disease characterized by low bone mineral density (BMD) and microstructural deterioration of the bone structure, leading to a consequent increase in fracture risk. Hip, spine, and wrist fractures caused by osteoporosis often lead to disorders that reduce the patient's quality of life and, in severe cases, increase the mortality risk. According to recent statistics from the International Osteoporosis Foundation, approximately one-third of women and one-fifth of men aged ≥50 years will experience an osteoporotic fracture. Osteoporosis is often not detected until fracture presentation and is hence considered a "silent epidemic" with a need for early diagnosis.

Raising awareness of osteoporosis increased treatment and treatment compliance rates, so awareness of osteoporosis may be the most effective strategy for prevention of osteoporotic fracture. In general, osteoporosis is diagnosed using BMD measured on central dual-energy X-ray absorptiometry (DXA), which is considered a reference standard test. However, the application of DXA is complex and expensive, and has limited availability for diagnosing the entire population. Other barriers to DXA screening are the absence of symptoms, which leads to low demand and reduced financial incentives for screening. As a result, nearly half of women covered by Medicare in the United States do not undergo DXA; even certain high-risk populations have low screening rates of ≤10%. Likewise, despite the reasonable prices and accessibility of DXA in South Korea, KNHANES IV 2008–2009 found that only 37.5% among the Korean general female population aged ≥50 years with osteoporosis were aware of their diagnosis and only 23.5% were under treatment for osteoporosis.

Recently, deep learning algorithms, especially convolutional neural network (CNN) architectures, have been widely recognized as an outperforming and reliable approach to identify clinically useful features directly from the medical images. Previous studies that used CNNs showed promising results in radiology, pathology, ophthalmology, surgery, and laboratory medicine. With the continuous improvement of the CNN architecture and the rapid increase in hardware computing power, CNNs have achieved human-level recognition performance. Nevertheless, only few CNN studies have been conducted for screening osteoporosis in radiographic examinations such as dental panoramic, lumbar spine, and hip radiographies. Studies have conducted osteoporosis screening in bones other than the vertebrae and femur on hand radiographs and panoramic radiographs. When using chest radiographs, readers often mention the presence of osteoporosis, but it is known that even if made by an experienced radiologist, the diagnosis of osteoporosis using chest radiographs is still unreliable. In general, the clues for screening osteoporosis on chest radiographs depends on the clavicles, ribs, and spines.

2.3. MACHINE LEARNING SOLUTIONS FOR OSTEOPOROSIS - A REVIEW [3]

The goal of **"**Teaching**"** machines to behave as if they are thinking, and thus appear intelligent, is commonly known as artificial intelligence (AI). Technically, AI comprises a wide variety of algorithmic tools intended to mimic human reasoning. Machine Learning (ML), a subfield of AI, involves the use of statistical methods to identify empirical patterns ("learn") from data. Specifically, ML uses the data as examples to guide learning process toward a given goal. Deep Learning (DL), a subfield of ML, provides multilayered ("deep") model architectures able to handle complex non-linear relationships between the input and output variables. With the Big Data Era and improvement of computing power, AI has revolutionized industries and, relatively recently, these approaches have been implemented in the field of medicine to find solutions to complex medical scenarios and multifactorial conditions. Significant successes have been found in diagnostic imaging, disease risk prediction, optimal treatment strategies decision-making, and prediction of molecular shape or activity.

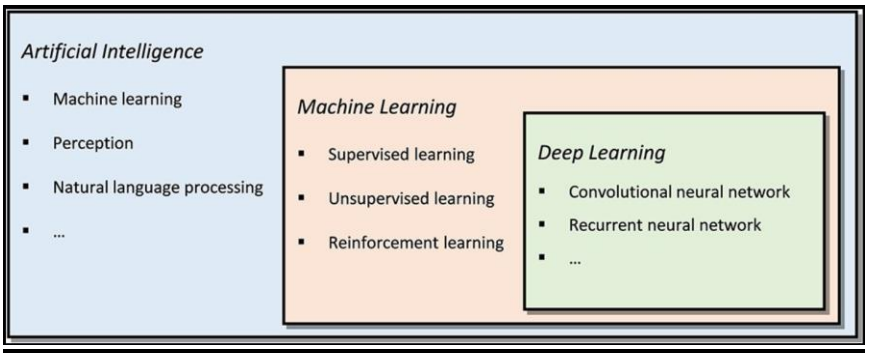

Fig 2. Hierarchical classification with examples of artificial intelligence, machine learning, and deep learning.

Osteoporosis is a complex disease in which the quantity and quality of bone are diminished, causing an increase in bone fragility. The clinical outcome of osteoporosis—fracture—occurs with an incidence of 8.9 million worldwide every year with considerable health, societal, and economic burden. Osteoporosis diagnosis is primarily based on bone mineral density (BMD) as assessed by dual-energy X-ray absorptiometry (DXA), but this does not capture the important contributions of clinical risk factors or other bone measures (eg, trabecular bone score, geometry). Preventing fractures is the main purpose in osteoporosis management. The main clinical risk factors for osteoporotic fractures and altered bone metabolism include older age, sex (with predisposition for women), ethnicity, heredity, previous fracture, malnutrition, alcohol consumption, current smoking, vitamin D deficiency, physical inactivity, various medications and medical disorders. FRAX is the most widely used and validated clinical tool for fracture prediction as based on many of the clinical risk factors referred to earlier combined with/without BMD. FRAX determines the 10-year probability of having a fracture by using classical statistical tools and is calibrated from country-specific data. AI/ML is finding novel

application in the deeper investigation of osteoporosis, including diagnosis and fracture prediction from biological testing, imaging, or clinical data.

2.4. PREDICTION OF OSTEOPOROSIS FROM SIMPLE HIP RADIOGRAPHY USING DEEP LEARNING ALGORITHM [4]

Osteoporosis is a common condition, especially in postmenopausal women; however, it often remains undetected until after fracture occurs. Early detection of osteoporosis is greatly important in preventing osteoporotic fractures. In the United States, the incidence of osteoporosis-related fractures is more than four times higher compared to that of stroke, heart attack, and breast cancer, and based on the meeting report of the World Health Organization (WHO), osteoporotic fractures account for more hospital bed-days than those diseases in several high-income countries. Hip fractures, one of the major osteoporotic fractures, are associated with limitations in ambulation, chronic pain and disability, loss of independence, and decreased quality of life, and 21%–30% of patients who have hip fracture die within 1 year.

To date, the gold standard for osteoporosis diagnosis is the estimation of bone mineral density (BMD) in the hip and lumbar spine using dual-energy X-ray absorptiometry (DXA). According to the WHO guidelines, BMD ≤ 2.5 standard deviations below the young adult mean (T-score ≤ -2.5) indicates osteoporosis, while a T-score at any site between -1.0 and -2.5 indicates low bone mass or osteopenia. Moreover, the US Preventive Services Task Force has recommended screening for osteoporosis with BMD testing to prevent osteoporotic fractures in women aged ≥ 65 years.

However, even though DXA is the gold standard of osteoporosis diagnosis, it could not be widely used as a screening tool for osteoporosis because of its high cost and limited availability in developing countries. To overcome these limitations, until now, great efforts have been made to develop a screening tool for osteoporosis. Quantitative ultrasonography is one of them, which has been developed as an alternative to DXA for screening osteoporosis. It is portable and more economical than DXA; however, it is insufficient to replace DXA as a screening tool for osteoporosis. Furthermore, there are also various clinical risk assessment tools that have been developed to predict osteoporosis, including Fracture Risk Assessment Tool (FRAX), QFracture algorithm, Garvan Fracture Risk Calculator, and the Osteoporosis Self-assessment Tool. These various risk assessment tools are easily accessible and useful; particularly, the FRAX calculator is a major achievement in terms of understanding and measuring fracture risk. However, a few limitations also exist, such as the lack of consideration of racial and ethnic group difference, especially those regarding body mass index and mortality rate. Therefore, an advanced screening tool for osteoporosis is still needed in clinical practice.

Recently, artificial intelligence (AI) has been used for various medical imaging interpretation fields. Moreover, several studies attempted to apply AI technology for the development of a screening tool for osteoporosis. Based on simple radiographic data, there were a few trials to predict osteoporosis using machine learning or deep learning algorithm. However, to the best of our knowledge, there were some limitations in the development of a real-world screening tool, such as inappropriate object detection results and extremely small sample size with methodological flaws.

2.5. OPPORTUNISTIC OSTEOPOROSIS SCREENING IN MULTI-DETECTOR CT IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORKS [5]

Osteoporosis is a common and frequently occurring disease in the aging population. About 200 million people suffer from osteoporosis and 89 million fractures occur worldwide every year. Osteoporosis is a disease of bone metabolism that shows a decrease in bone mineral density (BMD) and strength, which may lead to low back pain, disc degeneration, or an increased risk of fracture of the vertebral body. Hence, the early diagnosis of osteoporosis is very important to the progress of disease prevention.

Currently, evaluation methods for osteoporosis consist of the commonly used approaches, such as dual-energy X-ray absorptiometry (DXA), quantitative computed tomography (QCT), and quantitative ultrasound (QUS), and emerging imaging techniques, such as dual-layer spectral CT, H-MRS, and positron emission tomography (PET). BMD measurement is a reliable and ideal method for early diagnosis of osteoporosis. DXA is a commonly used tool for measuring spinal BMD. BMD measured by DXA is defined as the sum of cortical bone and cancellous bone, considering two-dimensional structures. However, DXA could not eliminate the influence of cortex, hyper osteogeny, and sclerosis on BMD measurement, which might underestimate the actual loss of bone mass. QCT is a recognized method for 3D bone density assessment. Several studies have shown that the detection rate of QCT on osteoporosis is significantly higher than that of DXA. But it requires calibration and standardized software, which means complex post-processing. And compared with DXA, the radiation dose of QCT is much higher. Thus, application of QCT as a screening technology is limited so far.

Millions of CT scans covering part or all of spine, are available from patients with other indications, such as urinary and / or digestive diseases, every year. These CT scans can be used for the opportunistic screening of osteoporosis, without additional exposure and substantial costs. Several literature studies had shown that conventional diagnostic CT scans were used to measure BMD by measuring directly the CT values of cancellous bone, leading to correlation coefficients ranging from 0.399 to 0.891. However, CT value not only depends on internal factors of vertebral body but also on external factors such as equipment, X-ray tube voltage, and CT device. Therefore, the CT values obtained from different devices need to be calibrated, which is why conventional CT scans in diagnosis of osteoporosis are limited.

Deep learning has been increasingly used in medical imaging analysis and even has entered the stage of rapid development. In terms of osteoporosis, several works on application of the deep learning technique have existed. Sangwoo et al combined machine learning and deep learning to predict patients with abnormal BMD by incorporating spine X-ray images. Bergman et al presented a deep learning method to compute the DXA BMD and T-Score from standard chest or abdomen CT scans. Pan et al developed a deep learning–based system to automatically measure BMD for opportunistic osteoporosis screening using low-dose chest CT scans obtained for lung cancer screening. However, in this system, only segmenting all vertebral bodies into three classes was used by a 3D CNN model, while isolating and labeling each individual vertebral body was then performed by conventional image processing algorithms. Using BMD values obtained with DXA as reference, Yasaka et al developed a deep learning

model to predict the bone mineral density of lumbar vertebrae from unenhanced abdominal CT images. This work only focused on BMD prediction, not provided vertebral body location. To the best of our knowledge, there are no studies regarding the application of deep learning on fully automated location of lumbar vertebral body and calculation of BMD similar to QCT value to date.

2.6. DEEP LEARNING OF LUMBAR SPINE X-RAY FOR OSTEOPENIA AND OSTEOPOROSIS SCREENING: A MULTI-CENTRE RETROSPECTIVE COHORT STUDY [6]

Osteoporosis and osteoporotic fractures have become global health issues of major concern with the growth in the aging population. By 2020, approximately 12.3 million individuals in the United States older than 50 years are expected to have osteoporosis. One in three women aged over 50 years will have an osteoporosis-related fracture. As a precursor of osteoporosis, osteopenia also deserves attention because most fractures in postmenopausal women occurred in those with osteopenia. Hence, screening of osteoporosis and osteopenia is clinically desirable for fracture prevention. The US Preventive Services Task Force (USPSTF) recommends that women aged ≥65 years should be routinely screened.

Central dual-energy X-ray absorptiometry (DXA) is globally accepted as the reference standard for diagnosing osteoporosis and osteopenia. However, the application of DXA is limited by its low availability, which typically requires patients to travel to a referral center. Other barriers to DXA screening include knowledge deficits and declining financial incentives for screening. As a result, nearly half of female Medicare beneficiaries in the United States do not undergo DXA testing, and certain high-risk populations have screening rates of <10%; while in China, only 4.3% women aged ≥50 years have undergone testing, particularly in rural areas, the rate is only 1.9%. The measurement of DXA assumes the presence of only bones and muscles, which would be inevitably influenced by fat. In addition, as a two-dimensional projection technique, it cannot fully consider bone geometry, size and microstructure. Accordingly, DXA is underutilized, and osteoporosis remains underdiagnosed. Safe and cost-effective alternatives to improve these conditions are needed. Conventional X-ray devices are widely available in almost any hospital worldwide, which carry potentially useful information about BMD. Retrieval of BMD data available on lumbar spine X-ray scans ordered for other indications requires no additional cost, patient time, or radiation exposure, and these data can be retrospectively acquired. It could, therefore, expand population screening effort for osteoporosis. However, it is a challenging task to assess BMD on lumbar spine X-ray images by inspection.

Recently, a deep learning technique, known as the deep convolutional neural network (DCNN), has gained significant ground in the field of computer vision. Deep learning takes raw image pixels and corresponding class labels from medical imaging data as inputs and automatically learns feature representation with multiple levels of abstraction. Continuous improvements of DCNN architectures coupled with a rapid increase in hardware computational power have enabled DCNN to achieve human-level performance in layer tasks, such as facial recognition, game playing, and natural language processing. Numerous early studies have also shown the promising results of DCNN used in a variety of medical imaging, including radiology, pathology, dermatology, and ophthalmology.

2.7. DEEP NEURAL NETWORKS FOR AUTOMATIC DETECTION OF OSTEOPOROTIC VERTEBRAL FRACTURES ON CT SCANS [7]

Osteoporosis, a chronic progressive bone disorder related to loss of bone density and quality, affects 10.2 million Americans and 56.2 million people worldwide. Weakened bone leads to fragility fractures that are associated with loss of independence and a decrease in the quality of life. If osteoporosis is detected early, it can be effectively treated to reduce future fractures and morbidity. However, before the onset of symptomatic fractures, osteoporosis is frequently silent, resulting in under-diagnosis and under-treatment.

Osteoporotic vertebral fracture or OVF, a marker of osteoporosis, is the most common type of osteoporotic fracture. The prevalence of OVF is high in older adults, reaching 40% by the age of 80. Nevertheless, under-reporting of incidental OVFs remains common with 84% of OVFs not reported in computed tomography (CT) exams in one study. The under-reporting of incidental OVFs in routine CT exams is attributed to radiologists' inattention to the sagittal views, the absence of clinical symptoms, and a lack of awareness regarding the clinical importance of asymptomatic OVFs.

There are new opportunistic approaches to screen for osteoporosis. These options are opportunistic because they rely on CT examinations performed for indications not related to the spine. As a result, the radiologist could be the first to suspect osteoporosis based on the imaging findings. Furthermore, these new screening approaches are efficient, because they do not require extra imaging time or radiation dose.

Recent advancement of machine learning allows automatic diagnosis of various conditions on radiology exams. Such automatic diagnosis has many benefits. For instance, radiologists no longer need to perform the tedious task of screening for incidental findings, and the saved time allows them to interact more with patients and health providers. Furthermore, these automatic diagnostic tools can address the lack of access to expert radiologists in rural, small, or poor communities. Particularly, embracing machine learning technology for detecting OVFs can improve early diagnosis of osteoporosis, initiate treatment, and predict future fragility fractures. As a result, a successful OVF detection system could potentially decrease the socio-economic burden of osteoporosis.

Previous work on automatic OVF detection relied on multiple and fragmented steps on each vertebra. These approaches were inefficient for this detection task because they required vertebra segmentation and calculations of height loss ratio on individual vertebral bodies. In this study, we developed and evaluated an automatic detection system for OVFs, based on an end-to-end deep learning model, which does not require multiple segmentation and analysis steps for each vertebra. In our proposed system, we leverage a convolutional neural network (CNN) to extract radiological features from chest, abdomen, and pelvis CT exams. The resulting sequences of features were then aggregated through a sequence classifier to predict the presence of a vertebral fracture on a CT scan.

2.8. DEEP-LEARNING-BASED DETECTION OF VERTEBRAL FRACTURE AND OSTEOPOROSIS USING LATERAL SPINE X-RAY RADIOGRAPHY [8]

Osteoporosis-related fragility fractures lead to high mortality and morbidity, particularly in the increasing population of older people, and impose a major health and economic burden worldwide. Thus, effective prevention of fragility fracture is among the most essential tasks in the health care of older adults. Despite remarkable improvements in pharmacological interventions, the risk of fragility fractures remains high. The low fracture-detection rate in high-risk subjects with osteoporosis and morphometric vertebral fracture remains a serious problem in clinical practice. Measuring areal bone mineral density (BMD) using dual-energy X-ray absorptiometry (DXA) is a cornerstone in the evaluation of bone strength; however, the limited availability of DXA contributes to low detection rates of osteoporosis. The presence of fragility fractures portends an increased risk of future fractures together with a low BMD value. However, the majority of vertebral fractures (a most common type) can be asymptomatic, which also contributes toward the low treatment rate of osteoporosis.

Machine-learning models capable of detecting osteoporosis or the presence of morphometric vertebral fractures may improve the referral rate of high-risk individuals to DXA testing or pharmacologic treatment initiation. Spine radiography is one of the most frequently performed imaging modality for various reasons. According to the International Society for Clinical Densitometry (ISCD) guidelines, lateral spine radiography should be accompanied by BMD measurement when evaluating the fracture risk in older adults. Because of the wide availability of centres that perform spine radiography and the substantial amount of information radiography provides (including bone density, spinal configuration, and soft-tissue information), spine radiographs can be the optimal sources for the development of machine-learning models that can support clinical decision-making on the basis of individual fracture risk.

Prior studies showed the feasibility of applying deep learning models on various types of X-ray images to detect osteoporosis, although most studies were limited to hip or wrist X-rays. Detection of prevalent vertebral fracture using deep learning models has been less intensively investigated using spine X-ray scans compared to DXA vertebral fracture assessment (VFA), computed tomography, or magnetic resonance images. A large, well-conducted study of 12,742 DXA VFA scans demonstrated that convolutional neural network could detect prevalent vertebral fracture that predicted clinical fracture outcomes. However, whether these results can be replicated in spine X-ray scans of various regions taken from different manufacturers and protocols remains to be tested. Two studies showed that deep learning models could detect prevalent vertebral fracture or osteoporosis on spine X-ray scans. However, these studies were limited by relatively small sample size of around 300 scans and the potential of simultaneous detection of prevalent vertebral fracture and presence of osteoporosis from a single spine X-ray scan has not been investigated yet. In addition, whether these findings would improve clinical indications for DXA testing remains unclear.

2.9. MULTI-LEVEL CLASSIFICATION TECHNIQUE FOR DIAGNOSING OSTEOPOROSIS AND OSTEOPENIA USING SEQUENTIAL DEEP LEARNING ALGORITHM [9]

There are many reasons why osteoporosis, which is also called as **"Metabolic Bone Disease"**, can go unnoticed for a long time and happen often. Because of this, bone cells micro-architectural degeneration significantly increases the risk of bone fragility as well as bone fractures, especially in the spine and femur, but also in other parts of the body and thereby patient's height gradually decreases (Jalodia 2021; Klibanski, 2001; Watts et al. 2008). As soon as possible, all of us need to get a precise test and need to undergo therapy that would help things not get worse. If any personnel is suffered with osteoporosis, a bone mineral density (BMD) test known as DEXA is used to measure BMD in order to confirm it. The DEXA testing needs to be taken in the lower back, femoral neck, and upper hip area. In the diagnostic testing, prognosis, as well as to treat osteoporosis disease, BMD tests are very important, and it will express BMD in grams of mineral per square centimeter scanned (g/cm2). The DEXA test gives BMD in two standards, they are the T-score and the Z-score, the T- score values of $-1, <-1$, and the values of 1–2.5, indicates that the personnel is affected with osteoporosis, as defined by the World Health Organization (WHO) (Deo 2015). It is used to show how much the patient's score can be above or below the normal value. The Z-score also shows the patient's BMD value in comparing with personnel having the same age and same sex. This is called the patient's "**Z score**". The mandibular cortical width that can be measured from digital dental radiographic images can also be used to find people who have low bone mass density (BMD) values (DPRs). The osteoporosis and osteopenia terms are used to describe a loss in bone density and a greater rate of bone breakdown than bone creation, resulting in porous bones. Decisions about a patient's medical care, such as whether or not they need medical treatment, should only be made by a doctor or other licensed health care practitioner. Whenever we have a question about a medical problem, we should always see a doctor or other healthcare professional. More information may be gleaned from physical examination data using sequential structures than with traditional categorization approaches. Different physical examination characteristics may be identified using Sequential procedures, which can also be utilized to learn about each feature's contribution to osteoporosis and osteopenia diagnosis categorization.

Recently, the 4th industrial revolution has seen a flurry of interest in artificial intelligence (AI). AI relies heavily on machine learning (ML), a fundamental component of which is employed in medical research. Prediction and classification are two of the most important functions of a machine learning system. Watson, the most prominent machine learning algorithm, has been used in the pharmaceutical industry to diagnose, analyses images, and optimize treatments (Kavitha et al. 2016). Medical professionals might benefit by using the above said algorithm for prediction and categorization of their diagnostics and treatment of patients. We devised a method for estimating the lumbar spine's T-score and classifying osteoporotic and non-osteoporotic vertebrae using HU of lumbar CT based on the prediction function of MLs. In the Materials and Methods section, we've outlined the steps involved in creating the data collection. In osteoporosis, density is reduced and a change in the micro-architecture structure of the bone appears, which decreases bone tolerance and increases the risk of fracture (Fine 2006). In osteoporosis, there is a progressive loss of bone mass, which is then used to build the model. Disruption of bones microarchitecture results in a decrease in BMD, and changes in bone protein concentration and diversity. Among the most common osteoporotic fractures are those

of the hip, spine, and wrist. To be considered osteoporotic, a fracture must occur in a place with low BMD and increase in frequency beyond the age of 50.

Osteoporotic fractures are a leading source of morbidity and death apart from the obvious physical consequences. In the United States, women have a 40% lifetime risk and males have a 13% lifetime chance of suffering a hip, spine, or forearm fracture after the age of 50. In Sweden, women make up 46% of the workforce, while males make up 22%. The BMD of Africans and Americans is 6% higher than that of Caucasians and Asians, putting them at greater risk. Every fifteen seconds, someone in the European Union has an osteoporotic fracture, resulting in a broken bone. As many as 75% of women with osteoporosis don't realize they have the condition.

It is possible for any individual get affected with both primary (idiopathic) and secondary osteoporosis. Primary osteoporosis, which affects women after menopause, is known as postmenopausal osteoporosis. Men might also get senile osteoporosis, which falls under this category. Anyone may develop secondary osteoporosis as a consequence of certain hormonal abnormalities and other chronic illnesses, such as glucocorticoids or other ailments that cause bone loss via multiple processes. When the condition is caused by steroids, it is known as osteoporosis produced by steroids (glucocorticoids). A fractured bone is often the first sign of osteoporosis, which is why the illness is referred to as "the silent crippler" since many individuals don't realize they have it until it's too late. However, early identification and treatment of osteoporosis may reduce a person's risk of fracture to an absolute minimum. Studies employing Artificial Intelligence algorithms have been utilized to determine whether or not a person has osteoporosis for these reasons.

2.10. OSTEOPOROSIS RISK PREDICTION USING MACHINE LEARNING AND CONVENTIONAL METHODS [10]

Fracture due to osteoporosis is one of the major factors of disability and death in elderly people. Osteoporosis is common in postmenopausal women but is asymptomatic until a fracture occurs. The World Health Organization (WHO) estimates that 30% of all postmenopausal women have osteoporosis, which is defined as bone mineral density (BMD) 2.5 standard deviations below the young healthy adult mean (T-score$\leq$-2.5). Dual X-ray absorptiometry (DEXA) of total hip, femoral neck, and lumbar spine is the most widely used tool for diagnosing osteoporosis. However, mass screening using DEXA is not widely recommended as it is a high-cost method of evaluating BMD. Therefore, selecting patients for DEXA is an important task for cost-effective screening for osteoporosis.

A number of epidemiological studies have developed clinical decision tools for osteoporosis risk assessment to select postmenopausal women for the measurement of BMD. The purpose of these clinical decision tools is to help estimate the risk for osteoporosis, not to diagnose osteoporosis. The osteoporosis self-assessment tool (OST) is one of the clinical decision tools, which is a simple formula based on age and body weight. Although OST uses only two factors

to predict osteoporosis risk, it has been shown to have good sensitivity with an appropriate cutoff value. However, the decision tool has the limitation of low accuracy for clinical use.

Machine learning is an area of artificial intelligence research which uses statistical methods for data classification. Several machine learning techniques have been applied in clinical settings to predict disease and have shown higher accuracy for diagnosis than classical methods. Support vector machines (SVM), random forests (RF), and artificial neural networks (ANN) have been widely used approaches in machine learning.

The SVM is based on mapping data to a higher dimensional space through a kernel function and choosing the maximum-margin hyper-plane that separates training data. RF grows many classification trees built from a random subset of predictors and bootstrap samples. ANN is comprised of several layers and connections which mimic biological neural networks to construct complex classifiers. Logistic regression (LR) is another machine learning technique. LR is the gold standard method for analyzing binary medical data because it provides not only a predictive result, but also yields additional information such as a diagnostic odds ratio.

3. METHODOLOGY

The methodology section outlines the step-by-step approach to achieve the best model and classification algorithm for the osteoporosis detection. In the context of osteoporosis detection using transfer learning techniques and various classification algorithms, the methodology encompasses data collection, model development, training, evaluation, and validation procedures.

3.1. BLOCK DIAGRAM

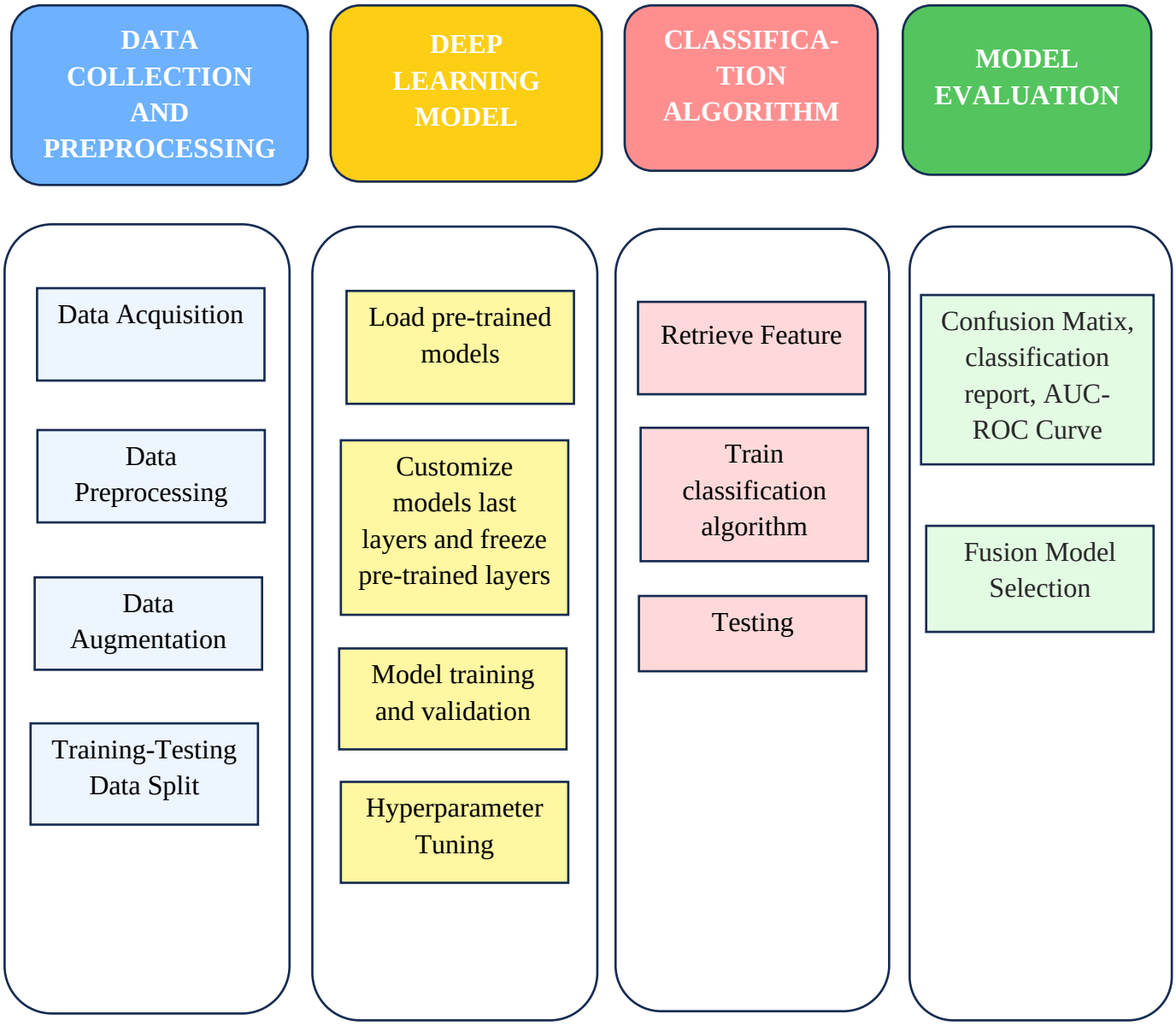

Fig 3. Block Diagram of finding best model for osteoporosis detection

3.2. DATA COLLECTION AND PREPROCESSING

3.2.1. Dataset Acquisition and Retrieval Procedure:

1. The dataset utilized for this research, titled "**Osteoporosis Knee X-ray Dataset**"**[11]** is accessed through the **Kaggle** platform, specifically curated by **StevePython**. This dataset is chosen for its relevance to osteoporosis detection and its availability on Kaggle, a reputable source of diverse datasets across various domains.
2. After locating the dataset on Kaggle, it is downloaded and subsequently uploaded to the researcher's Google Drive for convenient access during model development and experimentation. The decision to utilize Google Drive as a storage medium ensures seamless integration with **Google Colab**, a popular platform for machine learning and deep learning projects, facilitating efficient data handling and analysis.

3.2.2. Data Annotation and Labelling:

1. The **'flow_from_directory'** method automatically assigns labels to images based on their directory structure, ensuring accurate annotation.
2. Verification of class indices assigned by the generator is performed to validate the correctness of labelling.

3.2.3. Data Augmentation:

1. Augmentation techniques is applied for artificially expand the dataset, enhancing model generalization and robustness.
2. It Implement transformations such as rotation, scaling, and flipping to simulate variations in imaging conditions and patient positioning.
3. Augmentation techniques, including rotation, zoom, and flip, are applied within the **'ImageDataGenerator'** pipeline to introduce variations in image appearance and enhance dataset diversity.

3.2.4. Data Preprocessing:

1. Resize images to a standardized resolution (**224x224 pixels**) compatible with the input requirements of deep learning models.
2. Normalize pixel values to a common scale (e.g., [0, 1]) to mitigate variations in illumination and improve convergence during training.
3. Perform histogram equalization or contrast adjustment to enhance image contrast and accentuate relevant features.

3.2.5. Splitting the Dataset:

1. Divide the pre-processed dataset into training, validation, and testing sets, preserving the distribution of osteoporosis-positive and osteoporosis-negative samples.
2. The dataset is partitioned into training and testing sets using the **'train_test_split'** function, with careful consideration of **test size 20%** and **random state parameter 42** to control splitting and ensure reproducibility.

These meticulous steps ensure the acquisition, annotation, preprocessing, and quality assurance of the medical bone image dataset, laying a solid foundation for subsequent model development and evaluation.

3.3. TRANSFER LEARNING WITH EFFICIENTNETV2B0 AND VGG16

3.3.1. What Is Transfer Learning?

The reuse of a pre-trained model on a new problem is known as transfer learning in machine learning. A machine uses the knowledge learned from a prior assignment to increase prediction about a new task in transfer learning. You could, for example, use the information gained during training to distinguish beverages when training a classifier to predict whether an image contains cuisine.

The knowledge of an already trained machine learning model is transferred to a different but closely linked problem throughout transfer learning. For example, if you trained a simple classifier to predict whether an image contains a backpack, you could use the model's training knowledge to identify other objects such as sunglasses.

With transfer learning, we basically try to use what we've learned in one task to better understand the concepts in another. Weights are being automatically being shifted to a network performing "task A" from a network that performed new "task B."

To train a neural model from scratch, a lot of data is typically needed, but access to that data isn't always possible – this is when transfer learning comes in handy. Because the model has already been pre-trained, a good machine learning model can be generated with fairly little training data using transfer learning. This is especially useful in natural language processing, where huge labelled datasets require a lot of expert knowledge. Additionally, training time is decreased because building a deep neural network from the start of a complex task can take days or even weeks.

3.3.2. What is Keras?

Keras is a high-level, user-friendly API used for building and training neural networks. It is designed to be user-friendly, modular, and easy to extend. Keras allows you to build, train,

and deploy deep learning models with minimal code. It provides a **high-level API** that is intuitive and easy to use, making it ideal for beginners and experts alike.

Keras facilitates tasks like image classification, object detection, and video analysis through easy-to-implement convolutional neural networks (CNNs). This makes it ideal for applications from medical imaging diagnostics to automated manufacturing quality control. In NLP, Keras aids in building models for sentiment analysis, topic extraction, and machine translation. Its support for sequential data processing is essential for developing systems capable of summarizing texts or powering conversational agents. Keras is used for designing and training deep learning models efficiently and conveniently, handling tasks across various fields like image processing, natural language processing, and more due to its simplicity and robust API.

3.3.3. Deep Learning Model

In medical imaging, where data availability is often limited and model training from scratch may be impractical, transfer learning emerges as a powerful technique. It involves leveraging knowledge gained from pre-trained **deep learning models**, trained on large-scale datasets such as **ImageNet**, and adapting them to new, domain-specific tasks like osteoporosis detection. **ImageNet**, one of the most prominent image databases, contains millions of labelled images across thousands of categories. Models trained on ImageNet have learned to extract hierarchical features that are transferrable to a wide range of visual recognition tasks. By utilizing transfer learning with models pretrained on ImageNet, we can capitalize on the rich feature representations learned from diverse visual data. These pretrained models have already learned to detect basic visual patterns such as edges, textures, and shapes, which are highly relevant for medical image analysis tasks like osteoporosis detection. Fine-tuning these models on specific medical imaging datasets allows them to adapt their learned features to the nuances and intricacies of medical images, thereby enhancing their performance in tasks such as disease diagnosis and prognosis. For this project we have selected two models, **EfficientNetV2B0 Model** and **VGG16 Model.** We perform task, osteoporosis detection on both the model and on the accuracy based we select the final one for further updation.

3.3.4. EfficientNetV2B0: Deep Learning Architectures [12]

EfficientNetV2B0 represents a **state-of-the-art convolutional neural network architecture** that excels in both accuracy and efficiency. Its innovative design, incorporating compound scaling principles, allows it to achieve remarkable performance while minimizing computational demands. The architecture comprises a hierarchical structure of convolutional layers, progressively capturing intricate features at multiple scales. With its balance of effectiveness and efficiency, EfficientNetV2B0 serves as a compelling choice for medical imaging tasks, including osteoporosis detection. The **efficientnet-v2-b0** model, for example, is a variant pre-trained on the ImageNet dataset for image classification tasks.

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters
EfficientNetV2B0	29	78.7%	97.3%	7.2M

Table 1: EfficientNetV2B0 Model Accuracy, Size and Parameter

The total number of layers in EfficientNet-B0 the total is 237.

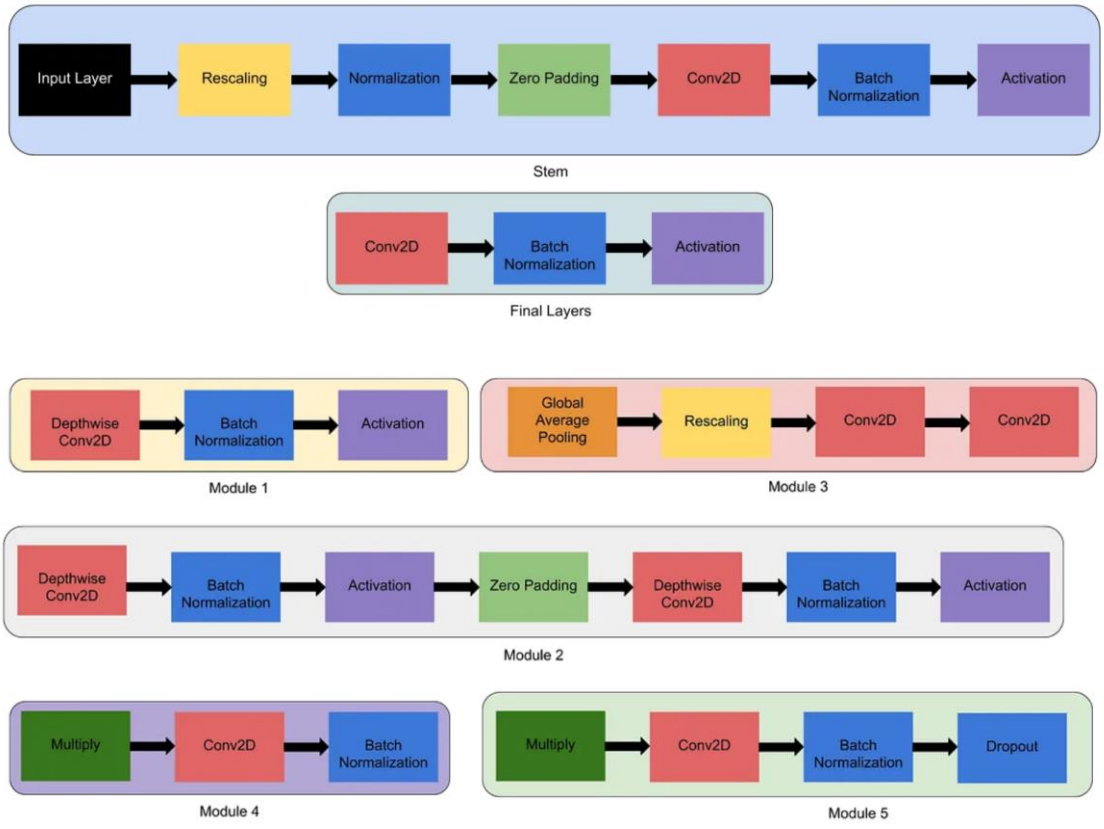

Fig 4. Layers in EfficientNetB2V0 Model

- **Module 1** – This is used as a starting point for the sub-blocks.
- **Module 2** – This is used as a starting point for the first sub-block of all the 7 main blocks except the 1st one.
- **Module 3 –** This is connected as a skip connection to all the sub-blocks.
- **Module 4 –** This is used for combining the skip connection in the first sub-blocks.
- **Module 5 –** Each sub-block is connected to its previous sub-block in a skip connection and they are combined using this module.

Till now we have specified everything that will be combined to create the Efficient Net models so let's get started.

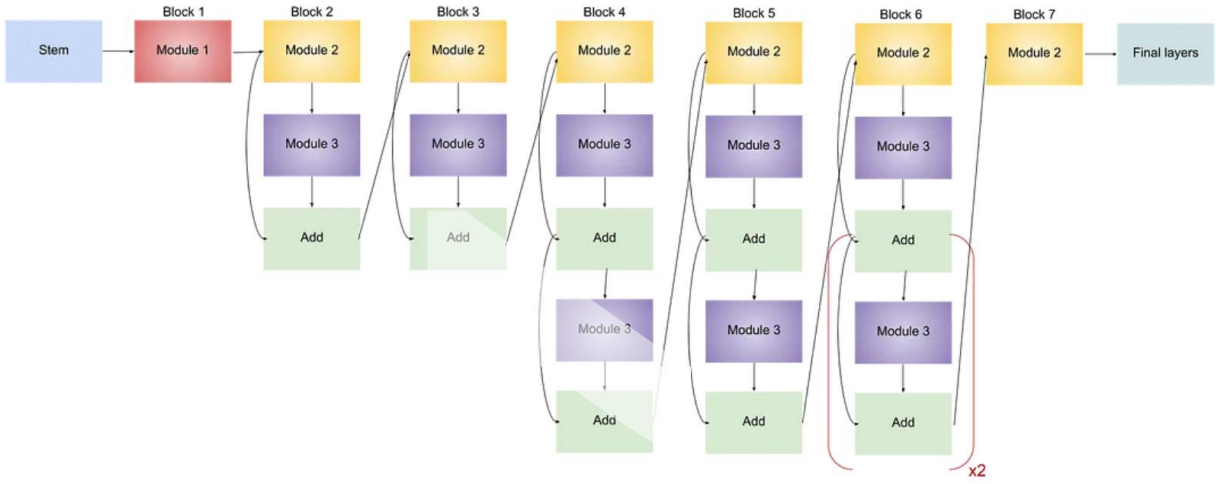

Fig 5. Architecture for Efficient Net V2 - BO. (x2 means that modules inside the bracket are repeated twice)

3.3.5. VGG16: Deep Learning Architectures [13]

The VGG-16 model is a **convolutional neural network (CNN)** architecture that was proposed by the **Visual Geometry Group (VGG)** at the **University of Oxford**. It is characterized by its depth, **consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers**. VGG-16 is renowned for its simplicity and effectiveness, as well as its ability to achieve strong performance on various computer vision tasks, including image classification and object recognition. The model's architecture features a stack of convolutional layers followed by max-pooling layers, with progressively increasing depth. This design enables the model to learn intricate hierarchical representations of visual features, leading to robust and accurate predictions. Despite its simplicity compared to more recent architectures, VGG-16 remains a popular choice for many deep learning applications due to its versatility and excellent performance.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition in computer vision where teams tackle tasks including object localization and image classification. VGG16, proposed by Karen Simonyan and Andrew Zisserman in 2014, achieved top ranks in both tasks, detecting objects from **200 classes** and classifying images into **1000 categories**.

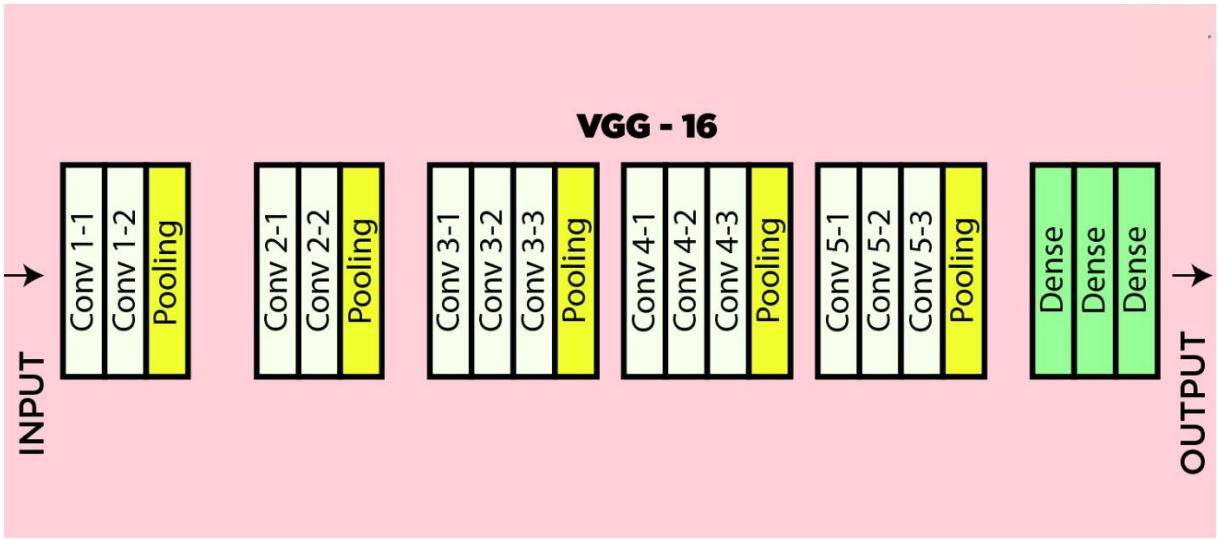

Fig 6. Architecture map for VGG16 Model

This model achieves **90.1% top-5** and **71.3% top-1** test accuracy on the ImageNet dataset which contains **14 million images** belonging to **1000 classes**.

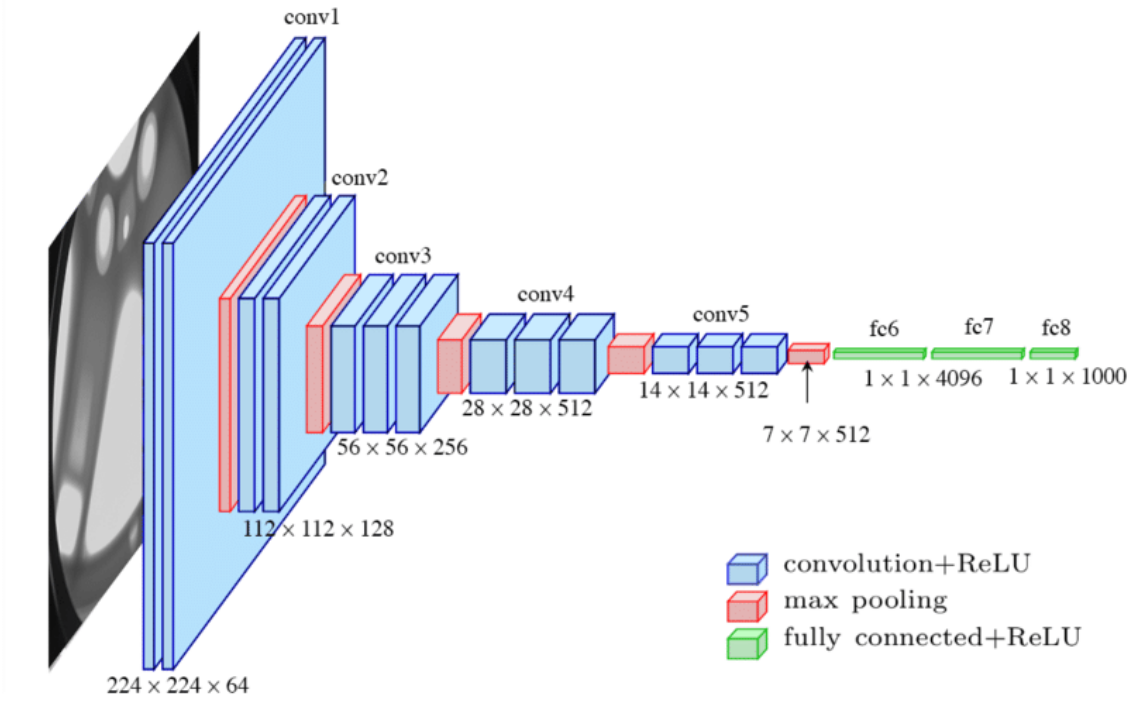

Fig 7. Architecture for VGG16 Model

3.3.6. Implementation in Code

The utilization of transfer learning with **EfficientNetV2B0** and **VGG16** unfolds in a systematic manner:

1. Model Loading and Customization:

- **i.** Pre-trained **EfficientNetV2B0** and **VGG16** models are loaded from the Keras library, preserving their learned weights and architectures.
- **ii.** The top classification layers of both models are discarded, leaving behind the convolutional backbone responsible for feature extraction.
- **iii.** Additional layers, such as dense layers, are appended to the truncated models to tailor them for osteoporosis detection. These added layers facilitate the translation of extracted features into meaningful diagnostic predictions.

2. Training and Evaluation:

- **i.** The customized models undergo training on the osteoporosis dataset, with training progress monitored and model performance assessed using various metrics.
- **ii.** To ensure model robustness and prevent overfitting, we employ model checkpointing, a crucial technique that periodically saves the model's weights and architecture during training. These checkpoints capture the model's state at different epochs, allowing us to resume training from the last saved point in case of interruptions and enabling model performance comparison across different epochs.
- **iii.** Performance evaluation encompasses not only accuracy but also sensitivity, specificity, and other relevant metrics, providing a comprehensive view of the models' diagnostic capabilities. By meticulously monitoring these metrics throughout the training process and leveraging model checkpoints, we can iteratively refine our models, fine-tune hyperparameters, and ultimately develop highly accurate and reliable diagnostic tools for osteoporosis detection.

3.4. CLASSIFICATION ALGORITHMS

To achieve optimal results, we strategically integrate classification algorithms with our deep learning models for osteoporosis detection. Model. The process involves passing the training data to the model, extracting features generated by the model from the training data, and then passing those features to the classifier to generate output. Here's how we leverage each algorithm to enhance the performance of our models:

3.4.1. Decision Trees and Random Forests:

1. We utilize decision trees as base learners within the ensemble framework of random forests.
2. Decision trees provide initial splits based on feature importance, while random forests aggregate multiple trees to improve robustness and reduce overfitting.
3. By combining the interpretability of decision trees with the ensemble learning of random forests, we achieve a balanced trade-off between accuracy and model complexity.

3.4.2. Support Vector Machines (SVM):

1. SVMs are employed as powerful classifiers to delineate complex decision boundaries between classes.
2. We harness SVMs' ability to maximize the margin between classes, ensuring effective separation of osteoporotic and non-osteoporotic bone images.
3. By fine-tuning SVM hyperparameters and kernel functions, we tailor the model to capture intricate patterns in the data, thereby enhancing diagnostic accuracy.

3.4.3. Naive Bayes and Logistic Regression:

1. Naive Bayes and logistic regression serve as baseline classifiers, providing simple yet effective models for osteoporosis detection.
2. We utilize their probabilistic nature to estimate the likelihood of osteoporosis based on input features, enabling intuitive interpretation of model predictions.
3. While Naive Bayes assumes feature independence, logistic regression accommodates more complex relationships through weighted combinations of features, offering flexibility in modelling.

3.4.4. AdaBoost:

1. AdaBoost is integrated into our framework to boost the performance of weak classifiers and adaptively focus on challenging samples.
2. By iteratively training multiple classifiers and assigning higher weights to misclassified instances, AdaBoost iteratively improves model accuracy and generalization.
3. The adaptive nature of AdaBoost ensures robustness against noise and outliers, enhancing the overall diagnostic capability of our models.

By strategically combining these classification algorithms with our deep learning architectures, we harness the complementary strengths of each approach to achieve superior performance in osteoporosis detection. The ensemble of classifiers offers enhanced robustness, interpretability, and generalization, paving the way for more accurate and reliable diagnostic solutions.

3.5. EVALUATION METRICS [14]

Evaluation metrics are essential tools for assessing the performance of machine learning models. In the context of osteoporosis detection, where the accuracy of diagnostic predictions is crucial, several key metrics are employed to gauge the effectiveness of the models. These metrics provide valuable insights into the model's ability to correctly classify individuals with and without osteoporosis.

1. Confusion Matrix

A **confusion matrix** is a matrix that measure summarizes the performance of a classification model on a set of test data. It is a means of displaying the number of accurate and inaccurate instances based on the model's predictions. It offers a thorough analysis of true positive, true negative, false positive, and false negative predictions, facilitating a more profound comprehension of a model's **recall, accuracy, precision,** and overall effectiveness in class distinction.

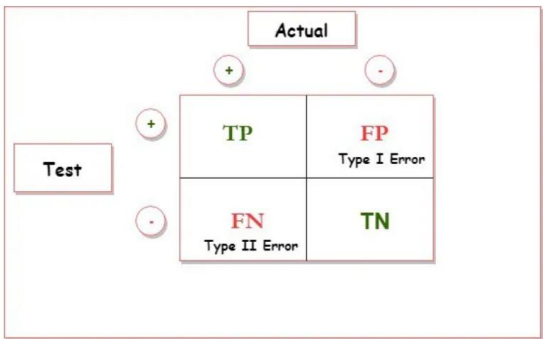

Fig 8. Confusion Matrix

- **True Positive**: The person has osteoporosis diseases and they actually have the disease.
- **True Negative**: The person not suffering from osteoporosis diseases and **actually** doesn't have the disease.
- **False Positives (FP):** Person has osteoporosis disease & they actually don't have the disease. (Also known as a "Type I error.")
- **False Negatives (FN):** Person does not have the osteoporosis disease & they actually have the kidney disease. (Also known as a "Type II error.").

2. Accuracy

Accuracy is used to measure the performance of the model. It is the ratio of Total correct instances to the total instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

3. Sensitivity OR Recall OR True Positive Rate

The proportion of actual positive cases that are correctly identified.

4. Specificity

The proportion of actual negative cases that are correctly identified. It measures the ability of a model to correctly identify negative instances. Specificity is also known as the True Negative Rate.

5. Precision OR Positive Predicated Value

Precision is a measure of how accurate a model's positive predictions are. It is defined as the ratio of true positive predictions to the total number of positive predictions made by the model.

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} = 1 - \mathrm{FPR}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$$

Fig 9. Recall, Specificity and precision formula

6. F1 Score

In some cases, data scientists and machine learning engineers try to obtain the best precision and recall simultaneously. F1 Score is the harmonic mean for precision and recall values. The formula for F1 score goes this way –

$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

The higher the F1 score more is the predictive power of the classification model. A score close to 1 means a perfect model, however, score close to 0 shows decrement in the model's predictive capability.

7. AUC-ROC (Area under ROC curve)

A **ROC curve (receiver operating characteristic curve)** is a graph showing the performance of a classification model at all classification thresholds, plotting between two parameter **True positive rate(TPR)** and **false-positive rate(FPR)**.

Points closest to the top left corner represents the optimal compromise b/w TPR and FPR.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$$

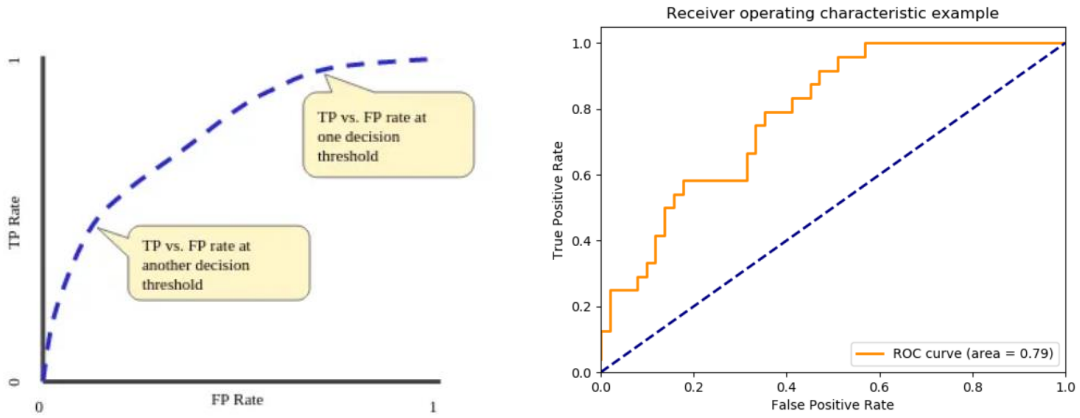

Fig 10. AUC-ROC Curve

AUC is an acronym for Area under the curve. It computes the entire 2D area under the ROC curve. AUC is a curve plotted between False Positive Rate Vs True Positive Rate at all different data points with a range of [0, 1]. Greater the value of AUC better the performance of the model.

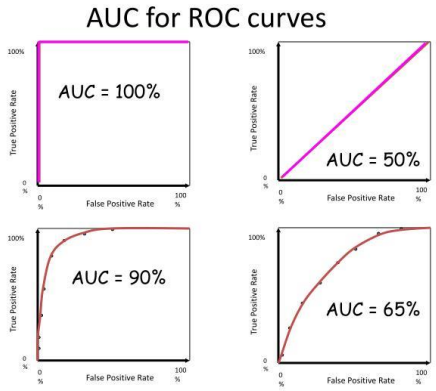

Fig 11. AUC Curves for different Values

4. EXPERIMENTAL RESULTS

4.1. DATASET DESCRIPTION

The osteoporosis dataset utilized in this study encompasses a collection of medical images crucial for detecting osteoporosis, a skeletal disorder characterized by weakened bones prone to fractures. These images are sourced from google drive; a directory located at **'/content/drive/MyDrive/osteoporosis dataset'**. Within this dataset, each image is categorized into one of two classes: **'Normal'**, indicating healthy bone structures, and **'Osteoporosis'**, indicating bones affected by the condition.

In total, the dataset comprises **372 images**, evenly distributed between the **'normal' and 'osteoporosis'** classes, with **186 images** representing each class. These images serve as the foundation for training and evaluating the deep learning models designed to detect osteoporosis.

To prepare the dataset for model training, preprocessing and augmentation techniques are applied using the **ImageDataGenerator**. This process involves resizing the images to a standard size of **(224,224) pixels** and organizing them into batches for efficient processing. Each batch consists of **20 images** paired with their corresponding labels, resulting in a total of **19 batches**.

During dataset generation, class indices are assigned to facilitate the mapping of class labels to numerical values. Specifically, the classes are indexed as follows: **'normal' (index 0)** and **'osteoporosis' (index 1)**. This indexing simplifies the process of representing class labels in a format suitable for model training.

Furthermore, to ensure the robustness and generalization of the trained models, the dataset is split into training and testing sets using a stratified approach. The training set comprises **297 images**, while the testing set contains **75 images**, maintaining the original class distribution. This stratified splitting strategy helps prevent bias and ensures that the models are exposed to diverse examples of both normal and osteoporotic bone structures during training and evaluation.

The training process is conducted over **20 epochs** with a **batch size of 16** that consist **0.8** specifies **80%** of training data aiming to iteratively update the model's parameters and optimize its performance on the training data. The **validation_split** parameter of **0.2** specifies that **20%** of the training data will be reserved for validation, allowing for the assessment of model performance on unseen data during training.

4.2. MODEL TRAINING

In our approach, we utilize pre-trained **EfficientNetV2B0** and **VGG16** models obtained from the Keras library, leveraging their learned weights and architectures. While these models are originally designed for predicting **1000** classes, we adapt them to suit our binary classification problem for osteoporosis detection. To tailor the models to our specific requirements, we modify their architectures by removing the top classification layers, including the extra layer from the **'avg_pool'** layer in the **EfficientNetV2B0** model and from the **'flatten'** layer in the **VGG16** model. By discarding these layers, we retain the convolutional backbone responsible for feature extraction while discarding unnecessary classification components.

Subsequently, we customize the models by adding our own custom layers to address the binary classification task. These additional layers, typically dense layers, are appended to the truncated models to facilitate osteoporosis detection. This process enables us to repurpose the powerful feature extraction capabilities of the pre-trained models while adapting them to the specific nuances of our diagnostic problem.

Model Summary of EfficientNetB2V0 model:

```
avg_pool (GlobalAveragePoo  (None, 1280)              0         ['top_activation[0][0]']
ling2D)

dense (Dense)               (None, 512)               655872    ['avg_pool[0][0]']

second_last (Dense)         (None, 64)                32832     ['dense[0][0]']

output (Dense)              (None, 2)                 130       ['second_last[0][0]']

==================================================================================================
Total params: 6608146 (25.21 MB)
Trainable params: 688834 (2.63 MB)
Non-trainable params: 5919312 (22.58 MB)
__________________________________________________________________________________________________
```

Model Summary of VGG16 model:

```
flatten (Flatten)           (None, 25088)             0

second_last (Dense)         (None, 512)               12845568

output (Dense)              (None, 2)                 1026

=================================================================
Total params: 27561282 (105.14 MB)
Trainable params: 12846594 (49.01 MB)
Non-trainable params: 14714688 (56.13 MB)
_________________________________________________________________
```

In the model training phase, we compile our custom model using the **Adam optimizer** and **binary cross-entropy loss function**, essential components for optimizing the model's performance in binary classification tasks. Additionally, we specify that the accuracy metric should be tracked during the training process to assess the model's classification performance.

To monitor the model's progress and ensure that we capture the best performing version, we implement a model checkpointing mechanism using the **ModelCheckpoint** callback from the Keras library. This callback allows us to define a filepath where the best model weights will be saved based on the validation accuracy metric. Specifically, the checkpoint is configured to save only the best model weights (those yielding the highest validation accuracy) and to do so in **HDF5** format for easy storage and retrieval.

During the model training process, the checkpoint monitors the validation accuracy and saves the corresponding model weights whenever an improvement is detected. This ensures that we retain the best-performing model configuration observed during training, even if subsequent epochs do not yield further improvements.

Finally, we execute the model training using the **fit() function**, specifying the training data **(x_train and y_train),** a **validation split of 20%** to assess model performance on unseen data, and a **batch size of 16** for efficient optimization. The defined **callbacks_list**, including the **ModelCheckpoint** callback, are passed to the **fit() function** to enable real-time monitoring and saving of the best model weights. The training process is conducted over **20 epochs in EfficientNetV2B0 model** and over **10 epochs in VGG16 model** to iteratively update the model's parameters and optimize its performance on the training data.

EfficientNetB2V0 Model

In the EfficientNetB2V0 model, the highest validation accuracy achieved during training is **0.8667**, indicating the model's performance in accurately classifying unseen validation data.

```
Epoch 8/20
15/15 [==============================] - ETA: 0s - loss: 0.6150 - accuracy: 0.7932
Epoch 8: val_accuracy improved from 0.83333 to 0.86667, saving model to /content/drive/MyDrive/Model Check Points/weights_best_EfficientNetV2B0.hdf5
15/15 [==============================] - 19s 1s/step - loss: 0.6150 - accuracy: 0.7932 - val_loss: 0.5967 - val_accuracy: 0.8667
```

Subsequently, when evaluating the model on the testing dataset, an accuracy of **0.7867** is attained, reflecting the model's ability to generalize well to new, unseen data.

```
3/3 [==============================] - 4s 1s/step - loss: 0.6263 - accuracy: 0.7867
Test Loss: 0.6263
Test Accuracy: 0.7867
```

Below are the visualizations depicting the training process, showcasing the evolution of training loss and accuracy over epochs.

The first graph illustrates the training loss, demonstrating how it changes over successive epochs. Meanwhile, the second graph presents the training accuracy, depicting how the model's accuracy evolves during training.

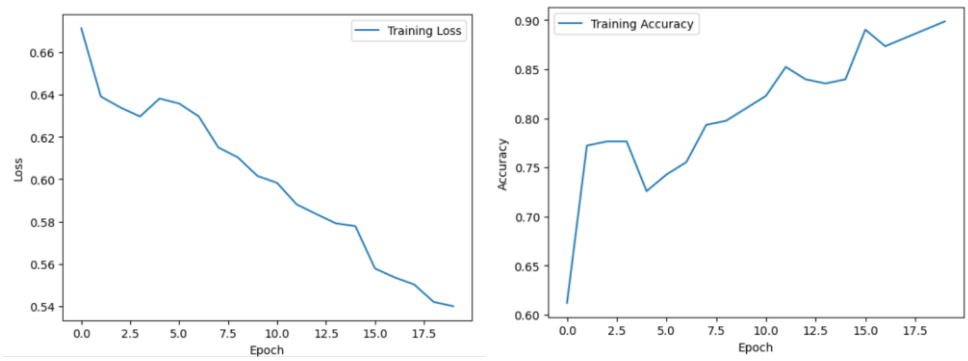

Fig 12. Training loss and training accuracy graph for EfficientV2B0 model

Below, we present the classification report generated by evaluating the model's predictions against the ground truth labels from the test dataset. The report includes precision, recall, F1-score, and support values for each class.

```
3/3 [==============================] - 8s 2s/step
              precision    recall  f1-score   support

           0       0.72      0.89      0.79        35
           1       0.88      0.70      0.78        40
```

Furthermore, we assess the model's performance using the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC), providing insights into the model's ability to discriminate between classes.

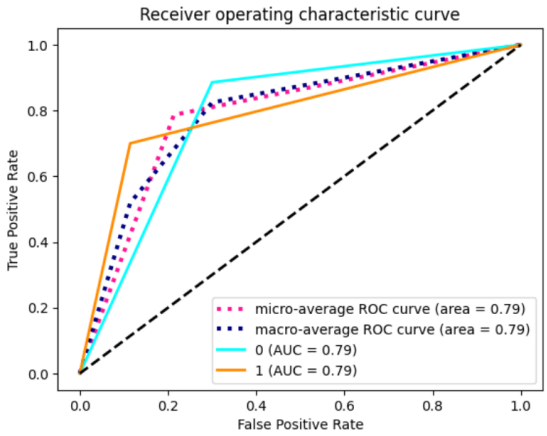

Fig 13. ROC Curve for EfficientV2B0 model

VGG16 Model

In the VGG16 model, the highest validation accuracy achieved during training is **0.8333**, indicating the model's performance in accurately classifying unseen validation data.

```
Epoch 2/10
15/15 [==============================] - ETA: 0s - loss: 3.8312 - accuracy: 0.8608
Epoch 2: val_accuracy improved from 0.78333 to 0.83333, saving model to /content/drive/MyDrive/Model Check Points/weights_best_VGG16.hdf5
15/15 [==============================] - 188s 13s/step - loss: 3.8312 - accuracy: 0.8608 - val_loss: 2.4665 - val_accuracy: 0.8333
```

Subsequently, when evaluating the model on the testing dataset, an accuracy of **0.7600** is attained, reflecting the model's ability to generalize well to new, unseen data.

```
3/3 [==============================] - 43s 13s/step - loss: 4.2532 - accuracy: 0.7600
Test Loss: 4.2532
Test Accuracy: 0.7600
```

Below are the visualizations depicting the training process, showcasing the evolution of training loss and accuracy over epochs.

The first graph illustrates the training loss, demonstrating how it changes over successive epochs. Meanwhile, the second graph presents the training accuracy, depicting how the model's accuracy evolves during training.

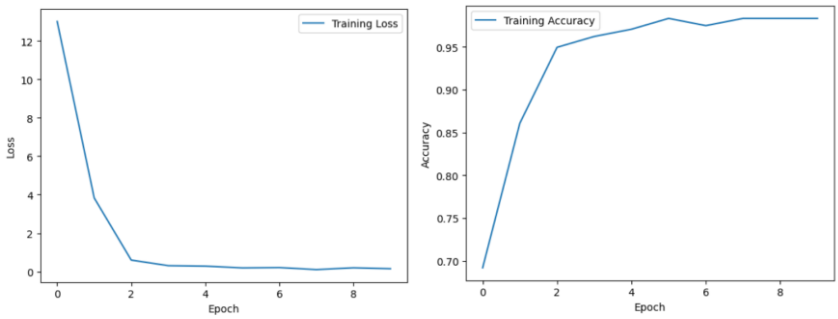

Fig 14. Training loss and training accuracy graph for VGG16 model

Below, we present the classification report generated by evaluating the model's predictions against the ground truth labels from the test dataset. The report includes precision, recall, F1-score, and support values for each class.

```
3/3 [==============================] - 43s 12s/step
              precision    recall  f1-score   support

           0       0.84      0.67      0.74        39
           1       0.70      0.86      0.78        36
```

Furthermore, we assess the model's performance using the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC), providing insights into the model's ability to discriminate between classes.

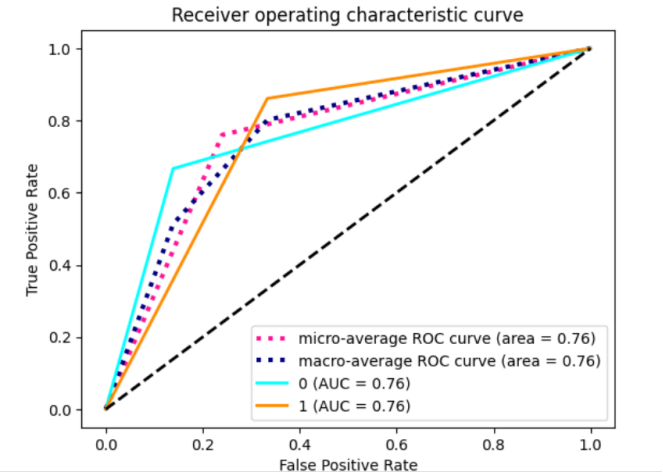

Fig 15. ROC Curve for VGG16 model

4.3. CLASSIFIER PERFORMANCE EVALUATION

We proceed by refining the model's predictive capability through strategic architectural modifications and subsequent training. Initially, we employ a transfer learning approach, leveraging pre-trained models such as **EfficientNetV2B0** and **VGG16**, both stripped of their top classification layers. This process allows us to harness the rich feature representations learned by these models and adapt them to our specific task of osteoporosis detection.

Subsequently, the extracted features from the modified models are flattened and utilized as input for various classification algorithms. This includes **Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes, Logistic Regression,** and **AdaBoost classifiers**. Each classifier is trained on the feature representations obtained from the pre-trained models, aiming to discern patterns indicative of osteoporosis across the dataset.

The choice of employing multiple classifiers is motivated by the desire to explore diverse modelling techniques and identify the most effective approach for accurate diagnosis. Additionally, the fusion of predictions from multiple classifiers holds promise in enhancing overall predictive performance, thereby contributing to the development of a robust diagnostic framework.

Furthermore, we meticulously evaluated the performance of each classifier using established metrics such as accuracy, precision, recall, and F1-score. These metrics provide valuable insights into the model's ability to correctly classify osteoporosis cases while minimizing false positives and false negatives. we constructed **receiver operating characteristic (ROC)** curves and calculate the corresponding **area under the curve (AUC)** for each classifier. This analysis offers a comprehensive assessment of the model's discriminatory power across varying thresholds, aiding in the selection of optimal classification strategies.

Classification Algorithms through EfficientNetV2B0 model

Classification Algorithm	**Accuracy**	**Value**	**Precision**	**Recall**	**F1-Score**	**Support**	**AUC**
Only EfficientNetB2V0 Model	0.7867	0	0.72	0.89	0.79	35	0.79
		1	0.88	0.70	078	40	0.79
Decision Tree	0.6133	0	0.57	0.69	0.62	35	0.62
		1	0.69	0.55	0.61	40	0.63
Random Forest	0.84	0	0.87	0.77	0.82	35	0.84
		1	0.82	0.90	0.86	40	0.84
Support Vector Machine	0.8134	0	0.76	0.89	0.82	35	0.82
		1	0.88	0.75	0.81	40	0.82
Naive Bayes	0.80	0	0.88	0.66	0.75	35	0.79
		1	0.76	0.93	0.83	40	0.79
Logistic Regression	0.7733	0	0.74	0.80	0.77	35	0.78
		1	0.81	0.75	0.78	40	0.78
AdaBoost	0.7866	0	0.76	0.80	0.78	35	0.79
		1	0.82	0.78	0.79	40	0.79

Table 2: Accuracy, matrix evaluation and other details for EfficientNetB2V0 Fusion Model

Classification Algorithms through VGG16 model

Classification Algorithm	**Accuracy**	**Value**	**Precision**	**Recall**	**F1-Score**	**Support**	**AUC**
Only VGG16 Model	0.76	0	0.84	0.67	0.74	39	0.76
		1	0.70	0.86	0.78	36	0.76
Decision Tree	0.7866	0	0.77	0.85	0.80	39	0.78
		1	0.84	0.72	0.78	36	0.80
Random Forest	0.8133	0	0.90	0.72	0.80	39	0.82
		1	0.75	0.92	0.83	36	0.82
Support Vector Machine	0.76	0	0.77	0.77	0.77	39	0.76
		1	0.75	0.75	0.75	36	0.76
Naive Bayes	0.8133	0	0.82	0.82	0.82	39	0.81
		1	0.81	0.81	0.81	36	0.81
Logistic Regression	0.8133	0	0.84	0.79	0.82	39	0.81
		1	0.79	0.83	0.81	36	0.81
AdaBoost	0.8	0	0.82	0.79	0.81	39	0.80
		1	0.78	0.81	0.79	36	0.80

Table 3: Accuracy, matrix evaluation and other details for VGG16 Fusion Model

5. DISCUSSION

5.1. INTERPRETATION OF RESULT

The experimental results offer compelling insights into the efficacy of our approach for osteoporosis detection:

1. **Transfer Learning Effectiveness:** The consistently high validation and testing accuracies achieved during model training underscore the potency of transfer learning in adapting pre-trained models to the task of medical image analysis. By leveraging knowledge gained from large-scale datasets, our models effectively learn informative representations from medical images, facilitating accurate osteoporosis detection.

2. **Algorithmic Variability:** The observed variation in performance across different classification algorithms highlights the significance of algorithm selection in model development. While some algorithms exhibit superior performance in certain aspects, such as accuracy or computational efficiency, others may excel in capturing nuanced patterns within the data. This underscores the importance of comprehensive algorithmic exploration to identify the most suitable approach for the given task.

3. **Random Forest Superiority:** Across both the **EfficientNetV2B0** and **VGG16** models, Random Forest consistently emerges as the top-performing algorithm. Its robustness in capturing complex relationships within the extracted features enables accurate classification of osteoporosis cases. The versatility and interpretability of Random Forest make it a compelling choice for medical image analysis tasks, offering clinicians valuable insights into disease detection and patient management.

4. **Comprehensive Evaluation:** The comprehensive evaluation of classification metrics, including **accuracy**, **precision**, **recall**, and **F1-score**, provides a holistic understanding of the developed framework's diagnostic capabilities. By examining multiple metrics, clinicians and researchers gain valuable insights into the model's performance across different aspects of osteoporosis detection, enabling informed decision-making and tailored intervention strategies.

In conclusion, our experimental results underscore the effectiveness of transfer learning in medical image analysis and highlight the importance of algorithmic selection in model development. The superior performance of Random Forest reaffirms its utility in accurate osteoporosis detection, offering clinicians a reliable tool for early disease diagnosis and intervention.

5.2. COMPARISON OF MODEL-ALGORITHM COMBINATIONS

The accuracy achieved by the **EfficientNetB2V0** model without classification algorithm is **78%**, while the **VGG16** model attained an accuracy of **76%** without classification algorithm during our experiments.

Upon incorporating classification algorithms as the final layer in both models, we observed notable enhancements in predictive performance for both **EfficientNetB2V0** and **VGG16** architectures.

In the case of the **EfficientNetB2V0** model, employing **Random Forest** as the classification algorithm yielded remarkable results, achieving an accuracy of **84%**. This approach demonstrated not only high accuracy but also exhibited commendable precision, recall, and F1-score metrics.

Similarly, within the **VGG16** model framework, the integration of **Random Forest** as the classification algorithm resulted in superior performance, achieving an accuracy of **81.33%**. Moreover, this combination exhibited robust precision metrics, indicating its effectiveness in distinguishing osteoporosis cases accurately.

A striking observation across both models was the consistent superiority of **Random Forest** over other classification algorithms, underscoring its efficacy in capturing intricate patterns within the feature representations extracted from medical images.

Furthermore, it's noteworthy that while individual model performances varied, the combination of **EfficientNetB2V0 architecture with Random Forest** consistently emerged as the **most effective model-algorithm fusion approach** for osteoporosis prediction, yielding the highest accuracy and precision metrics among all evaluated combinations.

5.3. STRENGTHS AND LIMITATIONS

Strengths:

1. **Effective Utilization of Transfer Learning:** Leveraging pre-trained models such as EfficientNetV2B0 and VGG16 facilitated efficient feature extraction from medical images, enhancing the diagnostic capabilities for osteoporosis detection.

2. **Robust Performance with Random Forest:** The integration of Random Forest as the classification algorithm demonstrated consistent and superior performance across both model architectures, achieving high accuracy and precision metrics.

3. **Comprehensive Evaluation:** By assessing multiple classification algorithms and conducting thorough performance evaluations, our approach provides a holistic view of the diagnostic framework's effectiveness, enabling informed decision-making in clinical settings.

4. **Generalizability:** The developed diagnostic model exhibits promising generalizability, as evidenced by its performance on independent testing datasets, suggesting its potential applicability across diverse patient populations and imaging modalities.

Limitations:

1. **Dependency on Quality of Training Data**: The effectiveness of the diagnostic model heavily relies on the quality and representativeness of the training data. Inadequate or biased datasets may lead to suboptimal performance and limited generalizability.

2. **Interpretability of Deep Learning Models**: While deep learning models offer powerful feature extraction capabilities, their inherent complexity often poses challenges in interpreting the underlying decision-making process, limiting the model's transparency and interpretability.

3. **Computational Resources**: Training and fine-tuning deep learning models, especially with large-scale datasets, demand significant computational resources and time, potentially limiting accessibility to smaller research groups or institutions with limited resources.

4. **Clinical Validation**: While the developed diagnostic model shows promise in early osteoporosis detection, further validation through clinical studies and integration into real-world healthcare settings is necessary to assess its clinical utility, safety, and efficacy in improving patient outcomes.

6. CONCLUSION

6.1. SUMMARY OF FINDINGS

6.1.1 Transfer Learning Enhances Diagnostic Performance:

Leveraging transfer learning with pre-trained models, EfficientNetV2B0 and VGG16, significantly improved the diagnostic performance for osteoporosis detection. Both models demonstrated competitive accuracy rates, highlighting the efficacy of transfer learning in medical image analysis.

6.1.2 Random Forest Emerges as the Top-performing Algorithm:

Among the evaluated classification algorithms, Random Forest consistently outperformed others, achieving the highest accuracy and precision metrics across both model architectures. This underscores the robustness and versatility of Random Forest in capturing complex patterns within medical image features.

6.1.3 Model-Algorithm Synergy:

The fusion of pre-trained models with appropriate classification algorithms, particularly Random Forest, yielded the most accurate diagnostic framework for osteoporosis detection. This synergistic approach capitalized on the strengths of both feature extraction and classification techniques, resulting in enhanced diagnostic accuracy and reliability.

6.1.4 Potential Clinical Implications:

The developed diagnostic framework holds promise for early osteoporosis detection, offering clinicians a valuable tool for risk assessment and fracture prevention strategies. However, further validation through clinical studies and integration into routine clinical practice is warranted to ascertain its real-world efficacy and impact on patient care.

6.2. FUTURE DIRECTIONS

6.2.1 Fine-tuning and Ensemble Methods:

Further exploration of fine-tuning techniques for pre-trained models could enhance model performance by adapting them more closely to the specific characteristics of osteoporosis imaging data. Additionally, ensemble methods, such as combining predictions from multiple models or algorithms, could be investigated to harness the strengths of different approaches and improve overall diagnostic accuracy.

6.2.2 Incorporation of Clinical Data:

Integrating additional clinical data, such as patient demographics, medical history, and biomarkers, could provide valuable contextual information to improve diagnostic accuracy and personalized risk assessment for osteoporosis. This multidimensional approach may enhance the predictive power of the diagnostic framework and enable more tailored interventions.

6.2.3 Multi-Modal Imaging Fusion:

Exploring the fusion of multiple imaging modalities, such as X-ray, MRI, and CT scans, could provide a more comprehensive assessment of bone health and fracture risk. By combining information from different modalities, synergistic insights into bone structure, density, and composition could be gained, leading to more robust and accurate diagnostic models.

6.2.4 Clinical Validation and Real-World Deployment:

Conducting rigorous clinical validation studies in diverse patient populations and healthcare settings is essential to assess the generalizability and real-world performance of the developed diagnostic framework. Collaborations with clinicians and healthcare institutions could facilitate the integration of the diagnostic tool into routine clinical practice, thereby enhancing patient care and fracture prevention efforts.

6.2.5 Explainable AI and Interpretability:

Incorporating techniques for explainable AI and model interpretability is crucial for enhancing trust and acceptance of the diagnostic framework among clinicians and stakeholders. Methods such as attention mechanisms, saliency maps, and feature attribution techniques can provide insights into the decision-making process of the models, enabling clinicians to better understand and interpret diagnostic predictions.

6.2.6 Longitudinal Monitoring and Risk Prediction:

Moving beyond binary classification, future research could focus on developing models for longitudinal monitoring of bone health and personalized risk prediction for osteoporosis-related fractures. By analysing temporal trends in imaging data and integrating longitudinal clinical data, predictive models could identify individuals at high risk of fracture and guide timely interventions to mitigate risk.

6.2.7 Integration with Telemedicine and Mobile Health Solutions:

Leveraging telemedicine platforms and mobile health applications could extend the reach of osteoporosis screening and monitoring to underserved populations and remote areas. User-friendly interfaces and automated decision support systems could facilitate seamless integration of the diagnostic tool into existing healthcare workflows, enhancing accessibility and early detection efforts.

By pursuing these future directions, we can advance the field of osteoporosis detection and management, ultimately improving patient outcomes and reducing the burden of osteoporotic fractures on global healthcare systems.

7. REFERENCES

[1] Iliou, T., Anagnostopoulos, C. N., & Anastassopoulos, G. (2014). Osteoporosis detection using machine learning techniques and feature selection. *Int. J. Artif. Intell. Tools*, *23*(05), 1450014.

[2] Jang, M., Kim, M., Bae, S. J., Lee, S. H., Koh, J. M., & Kim, N. (2020). Opportunistic osteoporosis screening using chest radiographs with deep learning: development and external validation with a cohort dataset. *Journal of Bone and Mineral Research*, *37*(2), 369-377.

[3] Smets, J., Shevroja, E., Hügle, T., Leslie, W. D., & Hans, D. (2020). Machine learning solutions for osteoporosis—a review. *Journal of bone and mineral research*, *36*(5), 833-851.

[4] Jang, R., Choi, J. H., Kim, N., Chang, J. S., Yoon, P. W., & Kim, C. H. (2021). Prediction of osteoporosis from simple hip radiography using deep learning algorithm. *Scientific reports*, *11*(1), 19997.

[5] Fang, Y., Li, W., Chen, X., Chen, K., Kang, H., Yu, P., ... & Li, S. (2021). Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. *European radiology*, *31*, 1831-1842.

[6] Zhang, B., Yu, K., Ning, Z., Wang, K., Dong, Y., Liu, X., ... & Zhang, S. (2020). Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone*, *140*, 115561.

[7] Tomita, N., Cheung, Y. Y., & Hassanpour, S. (2018). Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Computers in biology and medicine*, *98*, 8-15.

[8] Hong, N., Cho, S. W., Shin, S., Lee, S., Jang, S. A., Roh, S., ... & Kim, K. M. (2020). Deep-learning-based detection of vertebral fracture and osteoporosis using lateral spine X-ray radiography. *Journal of Bone and Mineral Research*, *38*(6), 887-895.

[9] Ramesh, T., & Santhi, V. (2024). Multi-level classification technique for diagnosing osteoporosis and osteopenia using sequential deep learning algorithm. *International Journal of System Assurance Engineering and Management*, *15*(1), 412-428.

[10] Kim, S. K., Yoo, T. K., & Kim, D. W. (2013, July). Osteoporosis risk prediction using machine learning and conventional methods. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 188-191). IEEE.

[11] Dataset used for Osteoporosis disease detection from Kaggle. Osteoporosis Knee X-ray Dataset (kaggle.com)

[12] EfficientNetB2V0 Model Images - Complete Architectural Details of all EfficientNet Models | by Vardan Agarwal | Towards Data Science

[13] VGG16 Model Images - VGG-16 | CNN model - GeeksforGeeks

[14] Evaluation Metrics Images - Complete Guide to Machine Learning Evaluation Metrics | by Shashwat Tiwari | Analytics Vidhya | Medium

- Osteoporosis Detection Code -

Fusion Model Using EfficientNetV2B0: EfficientNetV2B0_Fusion_model.ipynb - Colab (google.com)

Fusion Model Using VGG16 Model: VGG16_Fusion_model.ipynb - Colab (google.com)