Artificial Intelligence Methods in Osteoporosis Prediction Problem

ANNA HOVAKIMYAN[1], SIRANUSH SARGSYAN[1], TATEV HOVAKIMYAN[2],
ANI BADALYAN[1]
[1]Department of Programming and Information Technologies,
Yerevan State University,
1, Alek Manukyan St., Yerevan 0025,
ARMENIA

[2]AltMed Medical Center,
3/15 Hakob Hakobyan St., Yerevan 0033,
ARMENIA

Abstract: - Many sectors of human activity have implemented various solutions based on artificial intelligence methods. These solutions help significantly in decision-making tasks, especially when analyzing a large amount of relevant data is required beforehand. This paper discusses developing a computer system to assist doctors in diagnosing osteoporosis based on densitometric exam results. The system was developed using machine learning and trained on patient data obtained from densitometric examinations. The STRATOS device was used to collect data at AltMed Medical Center in Armenia. The goal of the system is to provide an accurate diagnosis of osteoporosis in patients while ensuring that the diagnosis is reliable and effective. During the system's development, we utilized three prominent machine learning models: Decision Tree, Random Forest, and SVM (Support Vector Machines). To enhance the accuracy and robustness of the system, these models were selected based on their effectiveness in solving complex classification problems. The developed system is equipped with advanced tools to detect potential diseases by exploring unidentified patterns and correlations among syndromes. The mentioned capability improves the diagnostic capabilities of the system. Achieving the medical goal requires early detection and accurate diagnosis. The AltMed Medical Center plans to utilize this system to provide medical professionals with support for informed decisions and improved patient care. The ability of the system to analyze complex medical data and reveal hidden insights makes it a valuable asset in the field.

Key-Words: - Osteoporosis, machine learning, classification task, osteoporosis diagnosis
Received: June 14, 2022. Revised: September 5, 2023. Accepted: September 29, 2023. Published: October 10, 2023.

1 Introduction

Osteoporosis is a common health issue that affects approximately 200 million people all around the world, according to the International Osteoporosis Foundation. It is more frequent in women than in men. Around one in every three women and two in every ten men aged 50 and above experience fractures related to osteoporosis, [1], [2], [3].

The impact of osteoporosis on individuals and healthcare systems is significant due to the associated fractures and their consequences. As the world's population ages, healthcare providers and policymakers face a growing challenge due to the projected rise in the prevalence of this disease.

The early detection and effective management of osteoporosis is crucial to mitigate its adverse impact on health and quality of life. The integration of machine learning can improve early detection rates and personalized treatment approaches, ultimately reducing the burden of osteoporosis on a global scale, [1], [4], [5]. Several factors can increase the risk of developing osteoporosis at any age, such as having other health conditions at the same time, having a genetic tendency for the disease, having insufficient levels of calcium and vitamin D in the blood, taking certain hormonal medications, smoking, drinking too much alcohol, and leading a sedentary lifestyle. It's important to diagnose osteoporosis as early and accurately as possible because there is a period of time when the disease may not show any symptoms, and early intervention can help prevent complications, [2], [3].

The conventional techniques for medical diagnosis, which rely on research findings, are facing challenges due to incomplete data and inaccuracies. Precise analysis of medical data is crucial for the accurate diagnosis, prognosis, and treatment of diseases. Artificial intelligence has emerged as a valuable tool for doctors to overcome these challenges, [4], [5].

Thanks to rapid advancements in computer technology, we now have access to a vast array of algorithms, models, and information technologies that ensure accurate and reliable data analysis. This has led to a significant decrease in incorrect diagnoses, resulting in more precise and dependable medical outcomes. The use of powerful computing and machine learning techniques in medical data analysis has truly revolutionized the field, [6]. Although medical technology has advanced, disease diagnosis remains a challenge for doctors, [7].

It is important to collect high-quality medical data and use appropriate methods to effectively address the issue of noisy, redundant, and incomplete data that can hinder predictive models in medical research.

To create intelligent physician assistants using artificial intelligence techniques, it is essential to have a substantial and relevant dataset. This data can be collected through laboratory tests or by utilizing hardware technologies like magnetic resonance imaging and computed tomography. Additionally, data from online sources such as Kaggle can also be utilized for this purpose, [8].

Obtaining precise medical information for certain diseases can be challenging, which can hinder the development of intelligent systems for disease prevention and diagnosis, [9]. Nevertheless, some computerized technologies that utilize medical diagnostic procedures are currently in use, particularly in countries such as Armenia.

For several years, the AltMed Medical Center in Armenia has utilized the STRATOS device for diagnosing osteoporosis, [10]. As a result, a significant amount of data has been collected, which necessitates the use of software tools for processing. Through the use of these tools, we can discover connections between research indicators that either confirm or disprove the existence of a disease that results from a combination of factors and irregular indicators. The identification of correlations and relationships among these characteristics has the potential to bring about a groundbreaking impact on the medical field. Doctors could benefit greatly from this, as it would help them prescribe more targeted treatments for specific diseases. For instance, in cases where hormonal therapy is used to treat thyroid gland issues, it's important to consider the risk of inducing osteoporosis in the patient.

The paper introduces a software system that uses machine learning to analyze densitometric measurements acquired from the STRATOS device at AltMed Medical Center, [10].

The newly developed software system can be a useful tool for physicians. It can precisely diagnose or dismiss the presence of osteoporosis in patients by examining the digital results of densitometric exams. Furthermore, it can detect patterns among survey indicators and offer general statistical data for further analysis and decision-making.

2 Machine Learning Models for Osteoporosis Prediction System

The system in question is based on machine learning, which is a subfield of artificial intelligence, [9], [11]. It involves creating models that can learn from data and use that knowledge to make predictions or informed decisions, [6], [11].

During the training process, a machine learning model adjusts its internal parameters to minimize the difference between its predictions and the actual outcomes. This continuous refinement enables the model to enhance its accuracy and effectiveness in making future predictions or decisions.

The process of machine learning involves various crucial steps that are necessary for achieving successful outcomes:

- Data Collection, Cleaning, Preprocessing, and Transformation.
- Data Labeling.
- Selecting the Appropriate Machine Learning Algorithm or Model.
- Model Training on Prepared Data.
- Evaluating the Model on Training and Testing Datasets.
- Model Tuning and Optimization.

By following these systematic steps, machine learning practitioners can develop robust and effective models for a wide range of real-world applications.

When it comes to analyzing data in densitometry studies, it’s a smart decision to consider using machine learning models like Decision Tree, Random Forest, and SVM (Support Vector Machine). Each of these models has its unique strengths that can be advantageous for different aspects of the analysis, as shown by research in sources, [11], [12], [13], [14].

3 Structure and Implementation of Osteoporosis Prediction System

The following sections will describe the structure of the system that has been developed.

3.1 Data Collecting and Preprocessing

The developed system uses data from densitometry and lab studies conducted with the STRATOS device. Each data sample contains about 130 features (Figure 1, Figure 2).

Preprocessing data is a critical step in analyzing data and building machine learning models. It involves several tasks, including cleaning and normalizing the data, encoding categorical variables, dividing the dataset into training and testing sets, and selecting important features for the model, [11], [15].

In cases where there is not enough data available for model tuning, artificial data generation techniques can be used to ensure high accuracy and reliable performance of the models, [15].

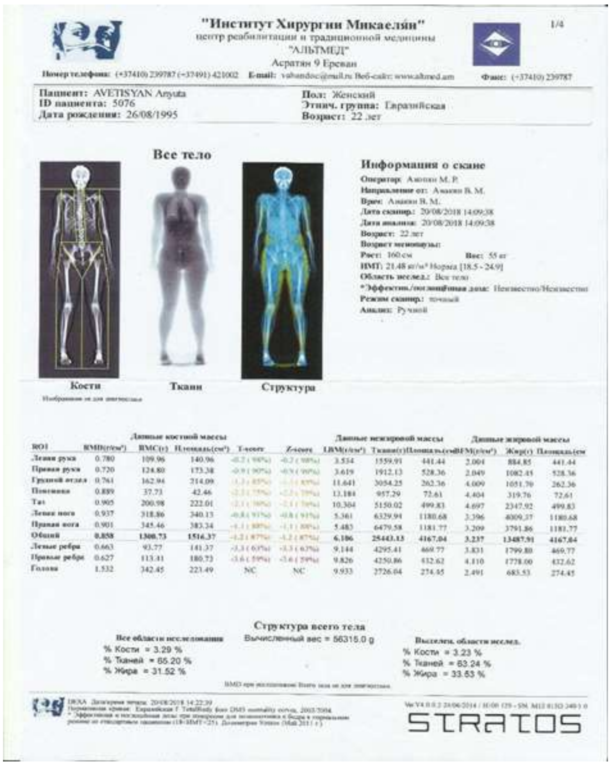
"Институт Хирургии Микаелян"
"АЛЬТМЕД"
1/4
Пациент: AVETISYAN Anyuta
ID пациента: 5076
Дата рождения: 26/08/1995
Пол: Женский
Возраст: 22 лет
Все тело
Информация о скане
Кости
Ткани
Структура
Структура всего тела
STRATOS

Fig. 1: Data received from the STRATOS device

	BS	BT	BU	BV	BW	BX	BY
1	general_lbm	left_ribs_lbm	right_ribs_lbm	head_lbm	left_hand_tissue	right_hand_tissue	chest_tissue
2	6.106	9.144	9.826	9.933	1559.91	1912.13	3054.25
3	6.906	10.793	11.019	10.55	2022.31	2397.43	3629.8
4	9.37	13.097	13.821	11.498	2840.9	3492.46	4643.87
5	8.108	12.132	12.321	11.768	3320.87	3894.27	4255.44
6	5.963	8.491	8.771	9.423	1386.88	1528.87	2405.18
7	6.075	9.254	9.434	9.01	1274.42	1694.85	2406.6
8	6.4	9.612	10.09	10.41	1529.87	1817.61	2551.71
9	6.106	9.144	9.826	9.933	1559.91	1912.13	3054.25
10	6.106	9.144	9.826	9.933	1559.91	1912.13	3054.25
11	9.37	13.097	13.821	11.498	2840.9	3492.46	4643.87
12	6.906	10.793	11.019	10.55	2022.31	2397.43	3629.8
13	6.906	10.793	11.019	10.55	2022.31	2397.43	3629.8
14	5.963	8.491	8.771	9.423	1386.88	1528.87	2405.18
15	6.4	9.612	10.09	10.41	1529.87	1817.61	2551.71

Fig. 2: Densitometric examination results

3.2 Labeling of Collected Data

Data labeling is an important step in supervised machine learning.

This step entails linking each data sample with its corresponding target output or label.

The machine learning model is trained using the labeled dataset to learn patterns and relationships between input features and target labels. After training, the model can predict labels for new data based on learned patterns, [11].

The labeling of data relies on indicators such as Z-score and T-score, as shown in Figure 3, [2], [3].

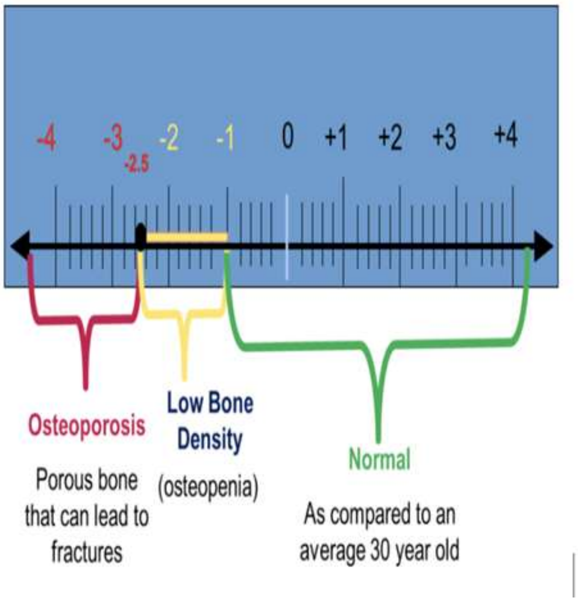

Fig. 3: Grades of osteoporosis on Z_score and T_score

3.3 Tuning and Evaluation of Machine Learning Models

The system uses machine learning models, including Decision Tree, Random Forest, and Support Vector Machine (SVM), which are implemented with Python's sci-kit-learn library. This library offers various machine learning algorithms and tools for analyzing data and developing models, [11], [12], [13], [14].

We have trained and evaluated three machine learning models - Decision Tree, Random Forest, and SVM - using a standard method of dividing the data into training and testing sets. The dataset was split into 70% for training and 30% for testing, ensuring accurate results.

The Decision Tree model achieved a training dataset accuracy score of 99.8%, while the Random Forest model scored 98.6%. On the testing dataset, the Decision Tree model scored 92.3% accuracy, compared to the Random Forest model's accuracy score of 95.8%.

The SVM classifier achieves a remarkable 89% accuracy using the Gaussian kernel, [14].

4 Correlation Data Analysis

The system analyzes patient data by focusing on age group distribution, gender, and Body Mass Index (BMI) to investigate the relationship between Z-score, T-score, and BMI characteristics and uncover potential connections and patterns. Medical practitioners can gain deeper insights into patients' health, identify hidden diseases early, and provide more targeted and effective care by studying the joint increase and decrease of Z-score, T-score, and BMI. The visual representation of connections in Figure 4 can help convey complex information and improve communication between healthcare professionals and patients.

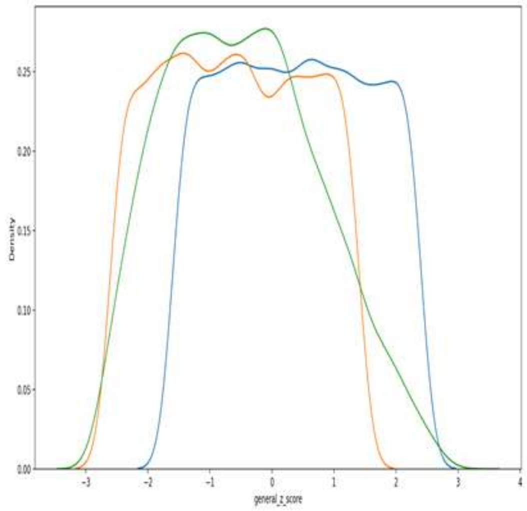

Fig. 4: Dependence of Z-score, T-score, and BMI

The system's ability to detect connections and correlations between different indicators from densitometric research is powerful in diagnosing and treating osteoporosis. Osteoporosis is a condition where bones become less dense and more prone to fracture, often without symptoms.

The cited examples provide valuable insights for diagnosing and treating osteoporosis. For instance, the BMI index and percentage of bone volume in the total body volume are independent of each other. Additionally, low bone density of the lumbar region is strongly correlated with low bone density of the wrist.

Using the connections and correlations it discovers, this system has the potential to improve osteoporosis management. It can help detect the condition early and assist in creating personalized treatment plans for people with different risk factors and bone health characteristics. In the end, these insights will help enhance patient care and outcomes when it comes to diagnosing and treating osteoporosis.

The correlation matrix shows the degree of correlation between indicators. Each point's color represents the correlation level (Figure 5). In a correlation matrix, each element represents the correlation coefficient between two variables. The matrix is usually presented in a square format. The correlation coefficient measures the strength and direction of the linear relationship between two variables, [16].

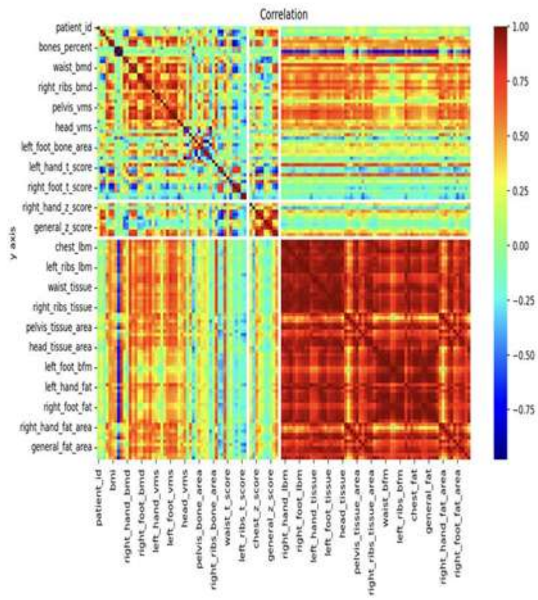

Fig. 5: Correlation between densitometric markers

In the matrix, the diagonal elements are always set to 1. This is because a variable has a perfect correlation with itself, resulting in a positive correlation of 1.0. In simpler terms, a variable always has the highest correlation with itself, creating a flawless linear relationship.

By studying the average masses of various bones per 1 cm² area (Figure 6), we can gain valuable insights into the variations in bone density among individuals. This study has revealed an important correlation between the densities of the lumbar and left ribs, with an 80% coincidence rate.

The diagram in Figure 7 shows that the densities of the sternum and lumbar bones have a lower coincidence rate of 20% in the examined individuals.

This finding suggests a weak correlation between bone densities in the sternum and lumbar regions, emphasizing the importance of separate evaluation for diagnosis and treatment planning.

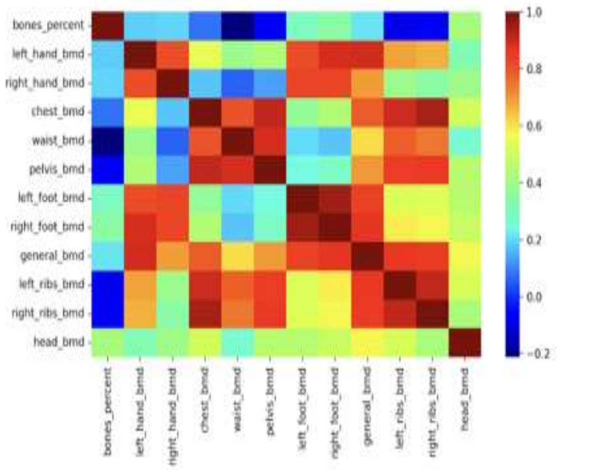

Fig. 6: Dependencies of bone density in different parts of the body

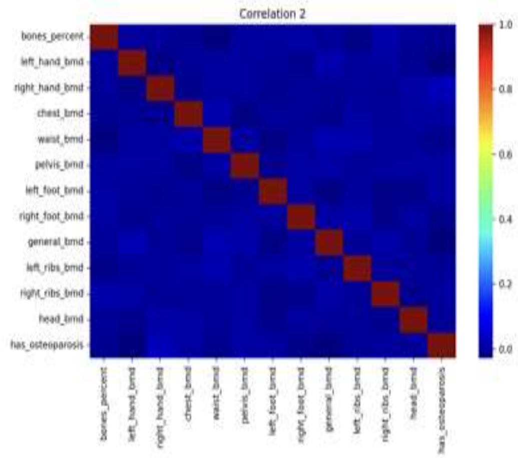

Fig. 7: Dependencies of bone density in different parts of the body

5 System's Interface

An easy-to-use interface has been developed for the intelligent diagnosis system, allowing doctors to efficiently utilize its functionality (Figure 8).

The doctor can easily download the patient's densitometry data and quickly receive the system's diagnosis for osteoporosis (Figure 9, Figure 10).

The doctor can access statistical analysis, calculate BMI, understand disease relationships, and train the system for improvement.

Fig. 8: Doctor's interface

Fig. 9: Disease prediction (negative result)

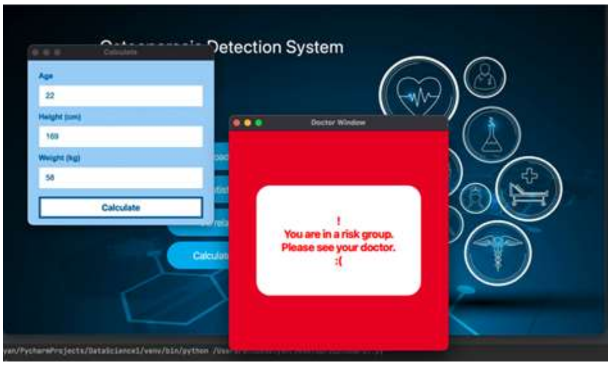

Fig. 10: Disease prediction (positive result)

6 Conclusion

A new technique for automating medical diagnoses using artificial intelligence and machine learning has been introduced in a recent paper. The software relies on three sturdy models and can accurately detect the presence of osteoporosis, given a sufficient amount and quality of medical examination results. This breakthrough has the potential to improve diagnostic effectiveness and enhance patient outcomes in osteoporosis detection.

The system is capable of discovering new connections and dependencies within research data. Its advanced data analysis capabilities can reveal previously unknown relationships and correlations, providing valuable insights and discovering novel patterns in medical research. This feature creates exciting opportunities for advancing scientific knowledge and improving decision-making processes in the field of medicine.

As the system continues to develop, it provides doctors with numerous valuable capabilities, such as:

- Identifying the underlying causes of disease, suggesting appropriate tests including the thyroid gland, diabetes, and genetic predisposition analysis for pediatric patients.
- Allowing doctors to seamlessly integrate and access additional test results from various devices, enabling more targeted and personalized treatment plans for patients.
- Utilizing the findings from investigations that concentrate on a specific body part and examining the connections that are uncovered, it can now be possible to anticipate the potential spread and severity of the illness in other parts of the body. This remarkable feat was once thought to be impossible, but thanks to the system's analysis, it is now attainable.

The system revolutionizes medical practice by offering advanced functionalities that enhance diagnostic accuracy and enable more precise and effective treatment strategies, ultimately leading to improved healthcare outcomes for patients.

References:

[1] Ann E. Kearns, *Mayo Clinic on Osteoporosis: Keep your bones strong and reduce your risk of fractures*, Mayo Clinic Press, 2021.

[2] Pauline M. Camacho, *Updates on Osteoporosis. An Issue on Endocrinology and Metabolic Clinics of North America*, Elsevier, 2021.

[3] "What is Osteoporosis?" by the National Osteoporosis Foundation, https://www.nof.org/ patients/what-is-osteoporosis/ (Accessed Date: 09/10/2023)

[4] "Osteoporosis Prevention: What You Need to Know" by WebMD, https://www.webmd.com/ osteoporosis/features/preventing-osteoporosis-tips (Accessed Date: on 09/10/2023)

[5] "Osteoporosis: Causes, Symptoms, and Treatments" by Medical News Today, https://www.medicalnewstoday.com/articles/155646 (Accessed Date: on 09/10/2023)

[6] Pranav Rajpurkar, Matthew P. Lungren, The Current and Future State of AI Interpretation of Medical Images, *The New England Journal of Medicine*, May 25, 2023, pp.1981-1990, DOI: 10.1056/NEJMra2301725

[7] Kyoung Hyup Nam, Il Seo, Dong Hwan Kim, Jae Il Lee, Byung Kwan Choi, In Ho Han, Machine Learning Model to Predict Osteoporotic Spine with Hounsfield Units on Lumbar Computed Tomography, *J Korean Neurosurg Soc.*, Vol.62, No.4, 2019, pp.442-449. DOI: 10.3340/jkns.2018.0178.

[8] Kaggle: Your Machine Learning and Data Science Community, www.kaggle.com (Accessed Date: on 09/10/2023)

[9] Mariya Yao, Adelyn Zhou, Marlene Jia, *Applied Artificial Intelligence: A Handbook for Business Leaders*, TOPBOTS, 2018.

[10] Altmed, *"AltMed" Medical Center: a unique combination of modern and traditional medicine methods!*, www.altmed.am

[11] Andreas C. Müller, Sarah Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, O'Reilly, 2017.

[12] Christopher M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2006.

[13] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly, 2017.

[14] Ingo Steinwart, Andreas Christmann, *Support Vector Machines*, Springer Science+Business Media, 2008.

[15] Prateek Joshi, *Artificial intelligence with Python*, Packt, 2019.

[16] Displayr, What is a Correlation Matrix?, https://medium.com/@displayr/what-is-a-correlation-matrix-b2196f728bc2 www.altmed.am (Accessed Date: on 09/10/2023)

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

- Anna Hovakimyan has proposed methods for solving the problem and was responsible for data collecting.

- Siranush Sargsyan organized the experiments and was responsible for results statistical analysis.
- Tatev Hovakimyan is an endocrinologist who treats osteoporosis.
- Ani Badalyan has implemented Machine Learning models in Python.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.