AI-Driven Classification of Bisphosphonate-Related Osteonecrosis of the Jaw (BRONJ) for Enhanced Clinical Management

Anita Petreska, Blagoj Ristevski IEEE Senior Member, Mirjana Markovska Arsovska, Nikola Rendevski

Abstract— **Bisphosphonate-related osteonecrosis of the jaw (BRONJ) is a rare yet serious condition that affects patients undergoing bisphosphonate therapy for osteoporosis and malignant bone diseases. Accurate classification of BRONJ stages is essential for early diagnosis and optimal treatment planning. Traditional diagnostic methods rely on subjective clinical assessments and radiographic analysis, which can lead to inconsistencies and delays in treatment. This study proposes an AI-driven classification model for BRONJ to address these challenges using machine learning (ML) and deep learning (DL) techniques. The research evaluates Support Vector Machines (SVM), Random Forest (RF), and Multilayer Perceptron (MLP) on a dataset containing demographic information, clinical symptoms, laboratory results, and patient medical history. To mitigate class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) and class merging strategies were applied. The optimized MLP model achieved 88.24% accuracy, improving generalization through regularization, dropout layers, and hyperparameter tuning. The results demonstrate the feasibility of ML and DL models for BRONJ classification, offering a more objective and automated approach to disease staging. The proposed model has the potential to reduce diagnostic variability, improve risk stratification, and assist clinicians in decision-making. By leveraging AI techniques, this study paves the way for more efficient and standardized BRONJ diagnosis, ultimately contributing to better patient outcomes and enhanced clinical workflows.**

Keywords— bisphosphonate-related osteonecrosis of the jaw; AI in healthcare; machine learning; deep learning; medical diagnosis; risk classification.

I. Introduction

The accurate classification of bisphosphonate-related osteonecrosis of the jaw (BRONJ) is essential for timely treatment and effective disease management. This condition is rare but serious, primarily affecting patients undergoing bisphosphonate therapy for osteoporosis or malignant bone diseases such as breast cancer, prostate cancer, and multiple myeloma [1]. Due to the risk of complications, it is crucial to establish preventive measures and standardized protocols for dental procedures to minimize the likelihood of BRONJ occurrence or progression [2], [3].

This study aims to predict the risk of bisphosphonate-related osteonecrosis of the jaws or its progression if it has already developed in patients who are on bisphosphonate therapy or have recently received it. Traditional diagnostic approaches rely on clinical assessments and radiographic analysis, but their subjectivity can lead to variability in diagnosis and delays in treatment [4], [5]. Advances in artificial intelligence (AI), particularly machine learning (ML), provide new opportunities for automated medical data analysis and improved clinical decision-making [6], [7]. Recent studies highlight AI's transformative role in medical diagnostics, enhancing early disease detection and pattern recognition across various conditions. For instance, Akhtar [21] discusses AI's application in diagnosing cancer, cardiovascular, and neurological disorders, while Mary [22] reviews AI's impact on diagnostic accuracy and speed. To enhance classification accuracy, this study develops a predictive model combining machine learning algorithms, including Support Vector Machines (SVM), Random Forest (RF), and a modified Multilayer Perceptron (MLP). These models are trained on a dataset containing demographic information, clinical symptoms, laboratory results, and patient medical history. Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) and class merging strategies are applied to address class imbalance and improve model stability. Furthermore, the MLP model is optimized using regularization, dropout layers, and hyperparameter tuning, resulting in improved generalization and an accuracy of 88.24%. This automated classification approach enables better risk stratification, reduced diagnostic variability, and enhanced clinical decision-making, potentially contributing to standardized early diagnosis and improved patient treatment.

The rest of the paper is structured as follows. Section II (Methodology) covers data analysis (EDA), class balancing techniques, ML model selection, and evaluation criteria. The next section details data preparation, class imbalance handling, and the training process of BRONJ classification models. The subsequent section discusses study limitations, including the limited dataset size, potential overfitting, and the lack of multimodal medical data. The last section summarizes the key findings and highlights the possible application of ML models in clinical decision support systems (CDSS).

II. Methodology

A. Exploratory Data Analysis (EDA)

To better understand the data and identify potential relationships between variables, we conducted an Exploratory Data Analysis (EDA), examining numerical and categorical

Anita Petreska is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, North Macedonia, e-mail: petreska.anita@uklo.edu.mk
Mirjana Markovska Arsovska, Faculty of medical sciences, University "Goce Delcev" Stip, e-mail:mirjana.arsovska@ugd.edu.mkcom
Blagoj Ristevski is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, Macedonia e-mail: blagoj.ristevski@uklo.edu.mk
Nikola Rendevski is with the Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" - Bitola, Macedonia e-mail: nikola.rendevski@uklo.edu.mk

attributes and their correlation with BRONJ stages [8]. The dataset consists of 59 patients with 13 features, categorized into numerical and categorical variables, which are valuable for analysing factors influencing the development and management of bisphosphonate-related osteonecrosis of the jaw (BRONJ). The age distribution of patients ranges from 39 to 88 years, with an average age of 67.8 years. The majority of patients belong to the 65–75 age group, suggesting that BRONJ predominantly affects older individuals. Gender analysis shows a significant predominance of females (61%) compared to males (39%), which is expected given that osteoporosis, one of the primary conditions treated with bisphosphonates, is much more common in postmenopausal women. The most frequent primary diagnosis in the dataset is osteoporosis (35.6%), followed by multiple myeloma (15.3%) and various types of cancers such as breast, prostate, uterine, and kidney cancer. Regarding treatment, the most commonly prescribed bisphosphonate is Zometa (zoledronic acid) (47.5%), followed by ibandronic acid (27.1%). A smaller percentage of patients (13.5%) received a combination or sequential bisphosphonate therapy. The average duration of bisphosphonate therapy is 3.82 years, with some patients receiving treatment for up to 15 years. The mandible is the most frequently affected region (62.7%), followed by the maxilla (27.1%), while both jawbones are affected in 10.2% of cases. The primary trigger for BRONJ development is tooth extraction (93.2%), whereas removable dentures (3.4%) and spontaneous cases (3.4%) are fewer common causes. The most frequently diagnosed BRONJ stage is "Risk" (32.2%), followed by Stage II (27.1%) and Stage I (20.3%), indicating that many cases are detected at an early stage, likely due to increased awareness and monitoring of patients undergoing bisphosphonate therapy.

B. Correlation and Statistical Insights

Cramér's V correlation matrix of categorical variables is presented in Figure 1, illustrating the relationships between different clinical factors associated with BRONJ. Of the 13 features, only 9 categorical variables are shown in Figure 1, as Cramér's V applies only to categorical data. Numerical variables like age and therapy duration were analysed separately using ANOVA and presented in Figure 2. To compute the correlation matrix, contingency tables were created for each pair of categorical variables. Cramér's V was calculated based on the chi-square statistic and normalized by the total sample size and the smaller number of categories, following the standard formula: $V = \text{sqrt}(\chi^2 / n\,(k - 1))$, where n represents the total number of cases and k the minimum number of levels across variables. This approach provided a robust, non-parametric assessment of associations among key clinical features, such as gender, disease type, therapy route, and BRONJ stage. A weak positive correlation (0.16) between age and bisphosphonate therapy duration suggests that age has little impact on how long patients remain on therapy. However, a moderate positive correlation (0.30) between age and therapy interruptions indicates that older patients tend to have longer treatment breaks, possibly due to medical conditions or physician recommendations. Similarly, a correlation of 0.28 between therapy duration and interruptions suggests that long-term therapy is more likely to involve pauses, likely as part of medical risk management. Although these correlations are not particularly strong, they highlight the relevance of therapy duration and patient age in treatment decisions and BRONJ risk assessment [9]. Cramér's V correlation analysis of categorical variables revealed key relationships in BRONJ progression. A strong correlation (0.92) between therapy type and administration route aligns with the standard practice of prescribing intravenous bisphosphonates (e.g., Zometa) for more severe conditions, while oral therapies (e.g., Alendronate) are mainly used for osteoporosis. Likewise, a correlation of 0.83 between disease type and administration route indicates that malignant conditions are more often treated with intravenous bisphosphonates. Gender and diagnosis showed a correlation of 0.76, reflecting the higher prevalence of osteoporosis in women and their increased likelihood of receiving bisphosphonate therapy. A moderate correlation (0.41) was found between therapy interruptions and BRONJ stages, suggesting that treatment pauses may contribute to disease progression. Similarly, therapy type and gender exhibited a correlation of 0.41, with women more likely to receive oral bisphosphonates, while men with malignancies are prescribed intravenous formulations. The correlation between BRONJ triggers and disease type (0.19) was weak, suggesting osteonecrosis can develop across different diagnoses. Additionally, there was no significant correlation (0.00) between affected jaw regions and therapy type, implying that BRONJ location depends more on dental interventions than bisphosphonate administration. These findings highlight the importance of personalized treatment planning, considering therapy type, demographic factors, and therapy interruptions.

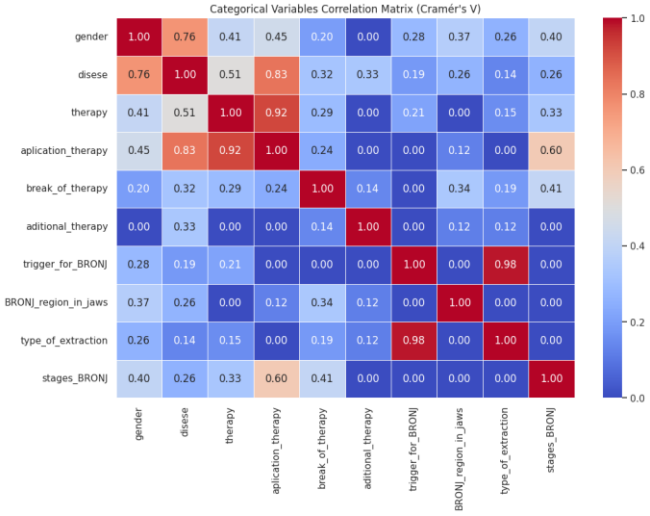

Figure 1: Correlation Matrix of Categorical Variables (Cramér's V)

The ANOVA F-value heatmap shown in Figure 2 provides valuable insights into the statistical relationship between categorical and numerical variables in the dataset. The strongest correlation was observed between therapy interruptions and the duration of these interruptions (F = 10.95), which is expected as patients who experience treatment breaks tend to have longer cumulative pauses in therapy. This suggests that therapy management strategies play a critical role in ensuring treatment continuity. Gender exhibited a notable correlation with bisphosphonate therapy duration (F = 4.66), indicating that male and female patients may have different treatment patterns. This is likely due to the higher prevalence

of osteoporosis in postmenopausal women, leading to longer therapy duration in female patients. A moderate correlation was found between additional therapies and therapy interruption duration (F = 3.90), suggesting that patients undergoing multiple treatments are more likely to experience treatment pauses. This could be attributed to the need for managing multiple comorbidities or adjusting therapy protocols. The association between disease type and therapy duration (F = 3.13) indicates that patients with different primary diagnoses, such as osteoporosis and malignancies, undergo varying bisphosphonate treatment lengths. Malignancy-related bisphosphonate treatments often involve more aggressive and extended therapy plans compared to those for osteoporosis [10]. Additionally, the mode of therapy administration influences therapy duration, with intravenous bisphosphonates typically prescribed for more severe conditions, leading to prolonged treatment periods. These findings highlight the importance of personalized treatment approaches to optimize patient care. Understanding how factors such as gender, disease type, and therapy administration impact treatment duration and interruptions can help refine clinical decision-making. By recognizing the influence of these variables, healthcare providers can develop more effective strategies to reduce therapy discontinuation, enhance adherence, and improve overall patient outcomes [11].

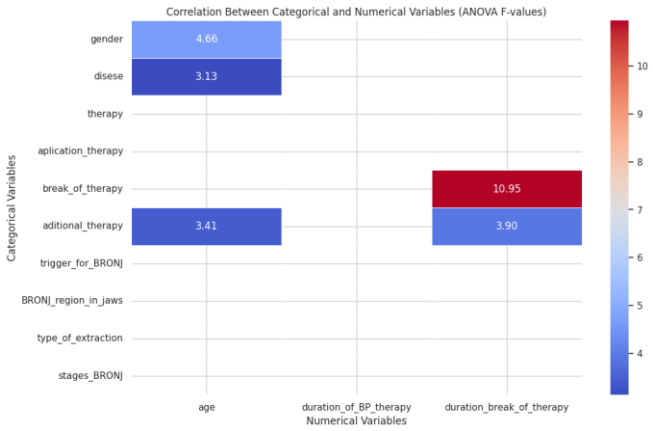

Figure 2: ANOVA Correlation: Categorical vs. Numerical Variables

C. *Profile Analysis by BRONJ Stages*

The progression of BRONJ across different stages reveals that both patient age and therapy duration influence disease severity. The average age increases from 66.3 years in the "Risk" phase to 73.3 years in Stage III, while the duration of bisphosphonate therapy extends from 2.72 years in Stage I to 4.63 years in Stage III. These trends suggest that prolonged bisphosphonate exposure may contribute to disease advancement, reinforcing the need for close monitoring to enable early detection and intervention. Therapy interruptions appear more frequently in advanced BRONJ stages, with 29.4% of Stage II and 36.3% of Stage III patients experiencing treatment pauses. While these interruptions may indicate greater disease severity, they do not necessarily cause BRONJ progression. Conversely, most patients in the "Risk" phase (84.2%) had uninterrupted therapy, suggesting that continuous bisphosphonate use does not always lead to worsening disease. Tooth extraction remains the most common BRONJ trigger across all stages, affecting over 90% of cases, reinforcing its role as a major risk factor. A small proportion of cases in Stage II (5.88%) were linked to removable dentures or spontaneous occurrences, emphasizing the importance of preventive dental assessments before and during bisphosphonate therapy. The mandible is the most frequently affected site, particularly in advanced stages, with 72.7% of Stage III patients exhibiting mandibular BRONJ. As the disease progresses, more cases involve both the mandible and maxilla, suggesting a pattern of increased severity and complications. These findings underscore the importance of individualized patient management, incorporating long-term monitoring and preventive dental care to minimize BRONJ progression.

D. *Feature Importance Analysis for BRONJ Progression*

The feature importance analysis (Figure 3) provides insights into the most influential factors driving BRONJ progression, as identified by the Random Forest model. This analysis highlights age, bisphosphonate therapy duration, underlying disease, and therapy interruptions as the primary contributors to disease advancement. Age (Importance = 0.178) and duration of bisphosphonate therapy (Importance = 0.169) emerged as the most critical predictors of BRONJ progression. Older patients and those with prolonged bisphosphonate exposure face a significantly higher risk of developing advanced disease stages. Research indicates that extended bisphosphonate use reduces bone regeneration and increases jawbone vulnerability, directly correlating with a higher BRONJ risk. Underlying disease type (Importance = 0.131) and therapy type (Importance = 0.079) also play a significant role. Patients with malignancies, such as multiple myeloma and various cancers, are more frequently treated with intravenous bisphosphonates (e.g., Zometa), which are associated with a higher BRONJ risk than oral therapies like Alendronate. Therapy interruptions (Importance = 0.057) and administration route (Importance = 0.088) have a moderate influence, indicating that the method of drug administration (oral vs. intravenous) and the duration of therapy pauses may impact BRONJ progression, though they are not the primary determinants. Other factors, including tooth extraction type, affected jaw region, and patient gender, demonstrated lower importance in disease development. This analysis underscores the significance of age, therapy duration, and underlying disease in BRONJ progression. Identifying high-risk patients early and tailoring treatment strategies based on these key factors can enhance disease management and improve patient outcomes.

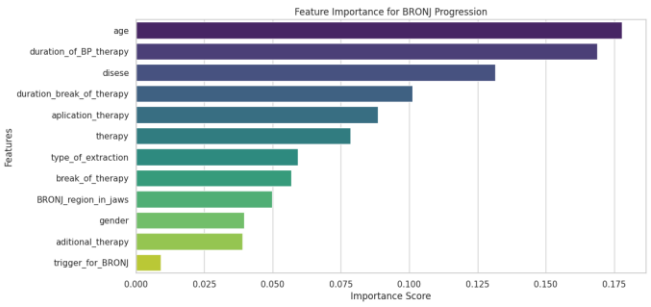

Figure 3: Feature Importance for BRONJ Progression

III. MODEL SELECTION & TRAINING

Predicting bisphosphonate-related osteonecrosis of the jaw (BRONJ) using machine learning and deep learning requires a well-prepared and balanced dataset. One of the key challenges in this process is class imbalance across different BRONJ stages, which can lead to unstable predictions and bias toward the most prevalent class. This issue arises because advanced stages of the disease are less common in clinical data, while most patients fall into early or risk stages. To improve model stability and accuracy, a class-merging strategy was applied, grouping similar clinical conditions to enhance data distribution and facilitate the classification process. Imbalanced class distributions pose several challenges in modelling. The more prevalent classes disproportionately influence model weights, while underrepresented ones are often ignored. This can lead to models predominantly predicting the dominant class rather than accurately reflecting the true data distribution. Additionally, models become unstable during training, particularly when significant disparities exist in sample sizes across classes. These effects can result in an artificially reduced error rate for dominant classes, while rarer cases remain underrepresented [12].

A. Data Balancing for BRONJ Prediction

To address this issue, a class-merging strategy was applied based on the clinical progression of the disease. "Stage 0" and "Stage I" were combined into a single group since Stage 0 represents an early phase without visible bone exposure, while Stage I includes exposed necrotic bone without signs of infection. Since both conditions represent early disease stages with similar treatments, merging them increased the sample size and improved class balance. Similarly, "Stage II" and "Stage III" were merged into a single category, as both represent advanced BRONJ stages with overlapping symptoms and treatment strategies. Patients in the "Risk" group remained in a separate category, as their clinical condition differs from those with developed BRONJ. After implementing this strategy, class distribution improved, enhancing the dataset's balance. Classifying patients into three groups instead of five allowed for more detailed analyses without losing clinical significance. This resulted in a better-distributed dataset, reducing model bias and improving accuracy. The merging of classes stabilized the model and reduced the likelihood of misclassification due to imbalanced data. Additionally, it enabled greater precision in algorithm training, as each class had a sufficient number of examples for meaningful learning [13], [14]. Applying such balancing techniques is crucial when handling medical data, particularly for rare conditions like BRONJ. Although this strategy significantly improved model performance, further research into alternative methods for reducing class imbalance remains essential. Approaches such as synthetic data augmentation using the Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, and advanced feature engineering could further enhance class differentiation. Implementing these techniques would improve ML models for BRONJ prediction, aiding in early detection and better disease management [15].

B. Class Balancing Using Oversampling (SMOTE)

To reduce the impact of class imbalance and improve model accuracy in BRONJ prediction, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. This approach increases the number of samples in underrepresented classes by generating synthetic data rather than simply duplicating existing instances [16]. Before applying SMOTE, the dataset distribution was uneven, with 28 samples in Stage II, 19 in the Risk stage, and only 12 in Stage I, which could lead to model instability and misclassification of early-stage BRONJ. After oversampling, all classes were balanced to 28 samples, ensuring a more equitable distribution and enhancing the model's ability to distinguish between different BRONJ stages. This balancing significantly improves model stability, reduces the risk of biased predictions, and enhances the learning of relevant features [17]. With a balanced dataset, overfitting is minimized, leading to better generalization and higher accuracy in clinical predictions. Future research may explore alternative techniques such as cost-sensitive learning and hybrid oversampling methods to further enhance the predictive power of the model while maintaining clinical relevance [18].

D. Machine Learning Models for BRONJ Classification

To predict and classify BRONJ stages, this study employs three machine learning models: Support Vector Machine (SVM), Random Forest (RF), and Multilayer Perceptron (MLP). The models were selected for their suitability with clinical tabular data: MLP models nonlinear patterns, RF is interpretable, and SVM works well with small samples. The models were chosen for their suitability with tabular data: SVM and RF perform well on small datasets, while MLP captures nonlinear patterns [19]. SVM was applied for its capability to construct optimal decision boundaries and handle high-dimensional data spaces. Different kernel functions enable better adaptation to data patterns, which is critical for distinguishing between various stages of the disease. RF, as an ensemble model, demonstrated high stability and interpretability by identifying key clinical features that significantly impact BRONJ progression. Its ability to aggregate results from multiple decision trees reduces bias and enhances model generalization [20]. MLP, as a deep neural network, was included for its ability to model complex nonlinear relationships between patient characteristics and disease stages. By implementing regularization and hyperparameter optimization, the model improved generalization without overfitting the data. The application of these three models allows for a comparative analysis of their effectiveness in predicting BRONJ stages. This study not only evaluates their performance but also provides insights into key predictors of the disease, contributing to improved clinical decision-making and a more personalized treatment approach [21].

E. Evaluation of Models for BRONJ Classification

The classification of bisphosphonate-related osteonecrosis of the jaw (BRONJ) requires a robust and well-generalized machine learning (ML) approach due to the complexity of disease progression and the overlapping clinical characteristics across different stages.

The ensemble approach of the Random Forest model contributed to improved generalization, while the deep

structure of the Multilayer Perceptron (MLP) enabled efficient recognition of complex patterns in the data. The initial iteration of the MLP model exhibited overfitting with 100% accuracy, but after modifications such as L2 regularization, dropout layers, and early stopping, its generalization significantly improved.

Table 1 presents the performance of the three ML models—Support Vector Machine (SVM), Random Forest (RF), and Modified Multilayer Perceptron (MLP)—evaluated using accuracy, precision, recall, and F1-score metrics. The Random Forest model achieved the highest accuracy of 94.12%, effectively classifying all BRONJ stages with high stability and generalization. The modified MLP model, after implementing regularization techniques to reduce overfitting, achieved an accuracy of 88.24%. On the other hand, SVM demonstrated slightly lower performance (82.35%), particularly in identifying patients in the early stages, indicating difficulties in handling nonlinear data distributions. The comparative analysis highlights that Random Forest provides the best balance between accuracy and interpretability, while the optimized MLP model offers greater flexibility in processing complex nonlinear relationships. SVM, though less efficient compared to the other models, remains applicable for smaller datasets with well-defined class boundaries. Future improvements should focus on hyperparameter optimization, ensemble learning techniques, and the development of hybrid ML-DL models to further enhance accuracy and generalization in BRONJ classification. Given the small sample size (n = 59), we used stratified k-fold cross-validation to maximize data use and preserve class distribution during evaluation.

TABLE I. PERFORMANCE METRICS OF CLASSIFICATION MODELS

MODEL	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
Support Vector Machine (SVM)	82.35%	88%	83%	83%
Random Forest (RF)	94.12%	94%	94%	94%
Modified Multilayer Perceptron (MLP)	88.24%	90%	89%	88%

IV. LIMITATIONS OF THE STUDY

Despite the promising results of the proposed machine learning models for BRONJ classification, certain limitations must be acknowledged to ensure their clinical applicability and reliability. The primary limitation is the size and representativeness of the dataset. Although the data used are real clinical records, they originate from a single clinical source, which is common in research on rare diseases like BRONJ. This factor may impact the generalizability of the models to broader patient populations. BRONJ is a rare condition, especially in advanced stages, which limits the availability of large, structured datasets. Nevertheless, the dataset used here consists of validated clinical cases that provide meaningful insights despite the small sample size. While balancing techniques such as SMOTE and class merging were applied, the dataset remains limited. SMOTE may introduce synthetic bias, so we applied stratified k-fold cross-validation to reduce its impact. SMOTE balances data but may add bias; stratified k-fold was used to reduce this. Future work will refine oversampling. Future work will explore advanced oversampling methods. Expanding the database with data from multiple centres would improve the robustness and external validity of the proposed models. Another key limitation is the potential risk of overfitting, particularly in deep learning models such as MLP. To improve MLP interpretability, future work will implement explainable AI techniques such as SHAP or LIME to identify feature contributions and build clinician trust in model outputs. Although L2 regularization, dropout layers, and early stopping were implemented to mitigate this issue, further improvements in hyperparameter tuning and cross-validation are required. Additionally, the exclusive reliance on structured tabular data excludes the potential influence of biomarkers and radiological data, which could enhance predictive accuracy. Integrating multimodal data, such as radiographic images and genomic markers, could significantly improve disease characterization and risk assessment [22]. A significant challenge is also the interpretability of deep learning models. Unlike Random Forest, which provides feature importance analysis, MLP lacks inherent transparency in decision-making, complicating its clinical implementation. Explainable AI (XAI) techniques, such as SHAP or LIME, can help provide insights into model predictions. Furthermore, this study does not incorporate longitudinal patient data, limiting its ability to track disease progression over time. Future research should implement temporal models, such as Recurrent Neural Networks (RNNs), to better predict BRONJ development over extended periods. Although PPV/NPV were not calculated due to limited sample size and missing outcome labels, future studies will focus on linking model outputs to treatment response benchmarks for clinical relevance.

A. Future Directions

To enhance predictive accuracy and clinical integration, future research should focus on expanding the dataset, incorporating multimodal data, and refining model architectures. Multi-institutional studies would improve model generalization, while the inclusion of radiographic and genomic data would enable more comprehensive risk assessments. Convolutional neural networks (CNNs) for radiographic image analysis and recurrent neural networks (RNNs) for longitudinal data tracking could further enhance predictive accuracy. Additionally, improving the interpretability of deep learning models is crucial through the implementation of explainable AI (XAI) techniques. This would provide a deeper understanding of the factors contributing to model decisions, reducing scepticism toward automated predictions. The introduction of SHAP, LIME, or other interpretability methods could facilitate the development of a transparent system that is more readily accepted in clinical practice. Model optimization through hyperparameter tuning and ensemble models could further boost classification accuracy. Techniques such as Bayesian optimization, genetic algorithms, and hybrid ML approaches (e.g., combining SVM, RF, and deep neural networks) may lead to superior results. Furthermore, the development of AI-powered CDSS would enable real-time BRONJ risk assessment and treatment recommendations, improving

clinical workflows and patient outcomes. Future research will include a cost-benefit analysis of model complexity versus clinical utility, comparing models in terms of training time, interpretability, and ease of integration into existing hospital systems. Ensuring compliance with data privacy standards (HIPAA, GDPR) is essential, with federated learning and privacy-preserving AI methods offering the potential to model sensitive clinical data without compromising security and confidentiality.

V. Conclusion

The evaluation of SVM, RF, and MLP demonstrated that ensemble-based methods like RF offer the best balance between accuracy and interpretability, while optimized deep learning models, such as MLP, effectively capture complex patterns in clinical data. To mitigate the challenges of BRONJ classification, this study employed data balancing techniques, including class merging and SMOTE, improving model robustness and ensuring a more equitable class distribution. Feature engineering, hyperparameter tuning, and regularization techniques further contributed to enhanced generalization, reducing the risk of overfitting. Beyond the BRONJ classification, this research underscores the transformative role of AI in modern healthcare. AI-driven models are revolutionizing medical diagnostics by enabling early disease detection, optimizing treatment strategies, and reducing diagnostic variability. The ability of AI to analyse vast amounts of clinical, genomic, and imaging data allows for more personalized and precise medical interventions. Furthermore, AI-powered systems can assist healthcare professionals in decision-making, streamline workflows, and improve patient management through predictive analytics and real-time risk assessment. We recognize the need for larger datasets to improve model generalizability. Future work will expand data via multi-centre collaboration and include multimodal inputs like imaging and biomarkers. Despite these advancements, certain limitations remain, including the dataset size, the lack of multimodal data integration, and the need for improved model interpretability. Future research should focus on expanding the dataset with multi-institutional clinical data, incorporating radiographic imaging and genomic biomarkers, and refining model architectures to enhance predictive accuracy. Additionally, the development of AI-powered CDSS could facilitate real-time BRONJ risk assessment and improve clinical workflows. This study demonstrates the feasibility of AI-based BRONJ classification, providing a foundation for future advancements in AI-driven medical diagnostics. By addressing the identified challenges and further integrating AI into healthcare systems, predictive models can be refined to support early detection, personalized treatment strategies, and improved patient outcomes. The integration of AI in healthcare is not only an innovation but a paradigm shift that holds the potential to significantly enhance clinical decision-making and patient care.

References

[1] Holen I, Coleman RE. Bisphosphonates as treatment of bone metastases. Current Pharmaceutical Design 2010; 16(11): 1262:1271.

[2] Manfredi M, Merigo E, Guidotti R, Meleti M, Vescovi P. Bisphosphonate- related osteonecrosis of the jaws: a case series of 25 patients affected by osteoporosis. International Journal of Oral and Maxillofacial Surgery 2011; 40(3): 277- 284.

[3] Khan Aa, Sandor GK, Dore E, Morrison AD, Alsahli M et al. Canadian consensus practice guidelines for bisphosphonate associated osteonecrosis of the jaw. Journal of Bone and Mineral Reaserch. 2007; 30(1): 3- 33.

[4] Da Woon K, Sung Min P. Prediction of medication-related osteonecrosis of the jaw (MRONJ) using automated machine learning in patients with osteoporosis associated with dental extraction and implantation: a retrospective study. J Korean Assoc Oral Maxillofac Surg 2023;49:135-141

[5] Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int 2020;51:248-57.

[6] Slavkovska, D., Petreska, A., Ristevski, B., Nikolovski, S. "Using Machine Learning Algorithms of Stroke Prediction." (2024): 461-471.

[7] Petreska, Anita, and Daniela Slavkovska. "Artificial Intelligence and Machine Learning Algorithms in Modern Cardiology." South East European Journal of Cardiology 5 (2024): 17-25.

[8] Petreska, A., Nikolovski, S., Novotni, G., Ristevski, B. "Analysis of Clinical, Genetic, and Demographic Data for Prediction of Alzheimer's Disease with Machine Learning." IEEE CoDIT 2024, pp. 569-574.

[9] Jamshidi, Alireza, Omid Moztarzadeh, Nasimeh Baghalipour, Soroush Najdaghi, Delaram Narimani Davani, Mohammad Behdad Jamshidi, and Lukas Hauer. "Using Machine Learning Techniques for Diagnosing and Analysing Medication-Related Osteonecrosis of the Jaw (MRONJ)." (2024).

[10] Kim, Jin-Woo, Jeong Yee, Sang-Hyeon Oh, Sun-Hyun Kim, Sun-Jong Kim, Jee-Eun Chung, and Hye-Sun Gwak. "Machine learning approaches for predicting bisphosphonate-related osteonecrosis in women with osteoporosis using VEGFA gene polymorphisms." *Journal of Personalized Medicine* 11, no. 6 (2021)

[11] Kim, Dong Wook, Hwiyoung Kim, Woong Nam, Hyung Jun Kim, and In-Ho Cha. "Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: a preliminary report." Bone 116 (2018): 207-214.

[12] Petreska, A., Ristevski, B., Slavkovska, D., Nikolovski, S., Spirov, P., Rendevski, N., Savoska, S. "Machine Learning Algorithms for Heart Disease Prognosis using IoMT Devices." (2023): 141-150.

[13] Salmi, Mabrouka, Dalia Atif, Diego Oliva, Ajith Abraham, and Sebastian Ventura. "Handling imbalanced medical datasets: review of a decade of research." Artificial Intelligence Review 57, no. 10 (2024): 273.

[14] Zhu, Junliang, Shaowei Pu, Jiaji He, Dongchao Su, Weijie Cai, Xueying Xu, and Hongbo Liu. "Processing imbalanced medical data at the data level with assisted-reproduction data as an example." BioData Mining 17, no. 1 (2024): 29.

[15] Elreedy, Dina, Amir F. Atiya, and Firuz Kamalov. "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning." Machine Learning 113, no. 7 (2024): 4903-4923.

[16] Özdemir, Akın, Kemal Polat, and Adi Alhudhaif. "Classification of imbalanced hyperspectral images using SMOTE-based deep learning methods." Expert Systems with Applications 178 (2021): 114986.

[17] Elor, Yotam, and Hadar Averbuch-Elor. "To SMOTE, or not to SMOTE?." *arXiv preprint arXiv:2201.08528* (2022).

[18] Edward, Jafhate, Marshima Mohd Rosli, and Ali Seman. "A new multi-class rebalancing framework for imbalance medical data." IEEE Access 11 (2023): 92857-92874.

[19] Puget, William, Olivier Mattelaer, and Fabio Maltoni. "A machine learning driven approach to the MLM merging algorithm."

[20] Petreska, Anita, and Blagoj Ristevski. "Artificial Intelligence in Intelligent Healthcare Systems–Opportunities and Challenges." In Serbian International Conference on Applied Artificial Intelligence, pp. 123-143. Cham: Springer Nature Switzerland, 2023.

[21] Akhtar, Zarif Bin. "Artificial intelligence within medical diagnostics: A multi-disease perspective." Artificial Intelligence in Health (2025): 5173.

[22] Umapathy, Vidhya Rekha, et al. "Perspective of artificial intelligence in disease diagnosis: a review of current and future endeavours in the medical field." Cureus 15.9 (2023).