Prediction of the 8 stages of osteoarthritis using deep learning.

Dr. Kumaragurubaran
School of Computer Science and Engineering
Rajalakshmi engineering College
Chennai, India
kumaragurubaran.t@rajalakshmi.edu.in

Abstract - This paper proposes a novel classification approach for knee osteoarthritis using an ensemble of CNNs that utilize TensorFlow. Osteoarthritis is a common degenerative joint disease often causing pain, limited mobility, and decreased quality of life, mainly in elderly individuals. Early and accurate diagnosis is crucial for timely intervention and efficient treatment planning. Our approach combines different CNN architectures into an ensemble model, which improves the feature extraction from knee radiographs more than the traditional single-CNN models. The ensemble model can capture a variety of patterns from images and could suppress overfitting because each CNN is trained on a different subset of the dataset. Advanced ensemble techniques, such as bagging and boosting, also enhance the model's accuracy and robustness. The proposed model was validated on a diverse set of knee radiographs that represent a spectrum of OA severities. Results show an improvement in classification accuracy, which indeed surpasses the traditional methods and, in some cases, even expert radiologists' assessments. Thus, this piece of research for the first time highlights the usability of CNN ensembles in medical imaging. The study underlines the importance of machine learning in furthering healthcare, as it provides the foundation for AI-supported systems to support clinical diagnosis and enable improvement in patient outcomes

Keywords- KneeOstheo arthiritis, Machine Learning, Deep Learning, Random Forest, Gradient Boosting, Bagging Regressor, RNN (Recurrent Neural Network), Stacking, Regressor, Meta-learner, One Hot Encoding, Normalization

I. INTRODUCTION

Knee OA is a common chronic degenerative joint disease affecting millions worldwide and significantly impairs quality of life. Characterized by cartilage breakdown, OA presents with symptoms such as pain, stiffness, and decreased mobility. Early diagnosis is what makes appropriate management possible. Without the previous diagnostic methods, which either rely on subjective examinations or radiographic imaging, variability in assessment and delays often occurred by their reliance on subjective assessments, which are prone to variability among practitioners and interpretation errors. These inconsistencies can delay diagnosis and compromise the development of optimal treatment plans.

Deep learning, especially CNNs, provides a more revolutionary answer. CNNs rely on vast numbers of medical image sets to recognize and classify patterns much more accurately and reproducibly, with considerably less dependence on subjective appraisals. This ability enables CNNs to detect knee OA earlier and much more reliably through its use of a standardized and reproducible diagnostic process. This shows that artificial intelligence may be a potential game-changer in medical diagnostics, particularly in musculoskeletal diseases. Incorporating CNN-based methodologies is a milestone toward enhancing the accuracy of diagnosis and patient care in knee OA treatment.

II. RELATED WORKS

The existing sEMG-PRS system has low generalizability for its use in practical applications. Here, we propose the idea of a SWRF algorithm to improve the long-term usability of the sEMG-PRS and user adaptability. First, we come up with the idea of WRF to abolish the drawback of the standard RF. By standard RF, there is a random issue while taking samples and features. Stacking is applied further to improve the generalization capability of WRF. RF is used as a base learner and WRF is employed as a meta-learning algorithm layer. The experimental results of SWRF is compared with some popular classification algorithms that are both online experiments and offline datasets. Offline experiments show that the average classification accuracy of SWRF is about 89.06%, which is higher than those of RF, WRF, long short-term memory (LSTM), and support vector machine (SVM). Online experiments reveal that the proposed SWRF surpasses the above algorithms concerning long-term usability and user adaptability. We consider that our method has good prospects for practical use in sEMG-PRS. The developed surface electromyography-based pattern recognition system (sEMG-PRS) has some principal drawbacks which prevent it from being implemented effectively in real-life applications. Surface electromyography-based pattern recognition systems (sEMG-PRS) face several critical limitations that significantly hinder their effectiveness and reliability in real-world practical applications. One of the major problems is the insufficient generalization capability because these systems tend to fail in demonstrating generalized performance by various user conditions, especially with dynamic usage environments. Such a limited capability of these systems is problematic in uncontrolled, real-world scenarios, where muscle signals and other environmental factors cause inconsistencies in outputs, thus largely hindering practical applicability. Another critical drawback is the imbalance in standard RF performance. RF is a very common used algorithm within sEMG-PRS. As it shows significant randomness in its sampling and feature selection algorithms, the models often suffer from a fluctuations in the classification accuracy levels, especially when handling high dimensional and complex data.

In addition to the above, a number of critical limitations accompany sEMG-PRS systems that considerably dampen their real-world application strength and reliability. A principal factor relates to generalizability: these systems are often unable to deliver consistent performance across various user conditions and dynamic usage environments. Such a limitation severely affects the usability of the system in uncontrolled, real-world situations, as variations in muscle signals and environmental factors may cause a lack of consistency in the output. Another significant issue that arises with standard RF performance is the imbalance in performance, the most commonly used algorithm in sEMG-PRS. Owing to the inherent nature of randomness to both sampling and selection of features in an RF model, these models generally show variability in their classification accuracy, mainly for high-dimensional and complex datasets.

An additional concern regarding the long-term usage of sEMG-PRS is its adaptability. Muscle signals in subjects may change over time due to effects like fatigue, physiological changes, or environmental factors. Current schemes do not adapt to these dynamic variations and, thus, their usability and reliability decrease. These adaptability challenges are exacerbated by the fact that existing systems were not able to realize the full potential of optimal classification accuracy, especially when compared to more advanced algorithms capable of exploiting complex patterns in sEMG data. The current solutions thus did not meet the required level of personalization, robustness, or consistent performance for practical, long-term use

To this end, we propose a novel methodology that can augment the generalization ability, adaptability, and overall performance of sEMG-based PRS systems. The proposed technique introduces a stacked weighted random forest (SWRF) algorithm, combining the benefits of WRF with the strengths of ensemble learning to overcome shortcomings in standard RF-based systems. WRF's algorithm helped to overcome these imbalances in performance by tackling the randomness involved with sampling and feature selection, making the whole process more stable and reliable for classification. For improving generalization, an approach known as stacking is proposed. Random forests become the base learners, while WRF forms the meta-learning layer. In this manner, the design becomes hierarchical and allows the SWRF algorithm to combine the benefits of each approach and increase its adaptability and robust performance.

The SWRF methodology is assessed rigorously by the method of using a combination of both offline datasets and online experiments. Offline testing proves that the average classification accuracy achieved by the SWRF is 89.06%, surpassing the performances of the conventional methods: standard RF, WRF, long short-term memory networks (LSTM), and support vector machines (SVM). The online experiments further validate the superiority of the SWRF algorithm for long-term use and adaptability of these traditional algorithms. In this context, the research proposed herein yields an important step forward in developing systems that can execute reliably in real-world settings and open the door for more practical and personalized applications of sEMG pattern recognition technology.

Another significant concern is the adaptability of the sEMG-PRS for long-term use. Users' muscle signals change due to a variety of factors that can occur over time, including fatigue, physiological variations, or even environmental influences. Current systems fail to accommodate these dynamic changes adequately, resulting in poor usability and lower reliability. The shortcomings in adaptability are further exacerbated by the limitations of existing systems in reaching the best possible classification accuracy, particularly in comparison with the performance of more sophisticated algorithms that can utilize intricate patterns that exist in sEMG signals. Thus, practical, long-term deployment requires solutions that better meet the requirements for more personalized, more robust, and predictable performance..

To address these challenges, we are proposing a novel methodology that can improve the generalizability, adaptability, and overall performance of sEMG-PRS. In particular, the proposed study introduces the SWRF algorithm, combining the strengths of WRF and ensemble learning techniques to overcome some of the main limitations of standard RF-based systems. The WRF algorithm controls performance imbalances by handling the randomness in sampling and feature selection, furthering stability and reliability in classification. To further enhance generalizability, a stacking approach is implemented. The base learners happen to be random forests, while WRF functions as the meta-learning layer. The hierarchical design lets the SWRF algorithm make use of both approaches for optimal adaptability and robust performance.

The proposed SWRF methodology is thoroughly tested by a combination of online experiments and offline testing. Offline experiments show that an average classification accuracy is obtained by SWRF to be 89.06%, surpassing the conventional methods like standard RF, WRF, LSTM, and SVM. Online experiments further test its superiority over these traditional algorithms in terms of the long term usability and adaptability. This work puts forward a significant step in systems toward reliable, real-world performance and lays the door open to more practical and personalized applications of sEMG pattern recognition technology reliable, real-world performance, paving the way for more practical and personalized applications of sEMG pattern recognition technology.

One major weakness is its lack of generalizability: sEMG-PRS is often not robustly reliable over diverse user conditions and varied usage environments. This potentially leads to inconsistent outputs, especially in real-world applications outside controlled laboratory settings. A second problem is that standard random forests (RF) used in such systems tend to have biased performance. Because the RF inherently operate based on random sampling and feature selection, these models tend to be unstable with respect to their classification accuracy, especially for complex, high-dimensional data. Moreover, such systems can be problematic in terms of long-term usability and adaptability because they fail to change when the users' muscle signals change over time, thus impairing the system's reliability and personalization.

Finally, the current methods lack optimal classification accuracy and are sometimes compared with more sophisticated approaches that take advantage of complex patterns in data and maintain performance. All of these limitations emphasize the need for a more robust, adaptive, high-performing solution for reliable use with sEMG-PRS in the long term.

III. Proposed Methodology

A. Problem Definition:

Surface electromyography-based pattern recognition systems generally have some significant limitations when it comes to performing reliably and effectively under practical, long-term scenarios. Limited generalizability is one of the primary problems with most such systems, which do not usually produce consistent results upon testing across various user conditions or in variable environments. Because of a lack of robustness on these lines, such systems are less suitable for real-world applications where dynamic and unpredictable factors frequently come into play. The use of standard random forests as the backbone classification algorithm is also one of the important factors causing the problem. Even though the RF technique is very popular, its randomness in both sampling and feature selection inherently causes imbalance in performance, leading to fluctuations in classification accuracies. Hence, it is, especially in handling complex high-dimensional datasets, further reduced the reliability of these systems.

Another important limitation of current sEMG-PRS is their limited ability to adapt flexibly to changes in users' muscle signals over time. Variations due to factors such as fatigue, physiological differences, or environmental changes often have a huge influence on the performance of such systems. Since such capacity to adapt to these evolving conditions does not exist, the long-term usability and reliability of such systems are compromised. These approaches are quite impractical and often result in less than ideal classification accuracy, so they cannot be compared to more advanced algorithms that can exploit complex patterns present in the data.

Therefore, there is a strong need for a robust, adaptable algorithm that has a high potential of overcoming these challenges. Such an algorithm should handle problems such as poorly balanced performance and accuracy inconsistency to guarantee the ability of the system to adapt dynamically to different user conditions over time. On top of improving classification accuracy, adaptability, and ensuring long-term usability, advancing the practical application of sEMG-PRS is primarily aimed at making these systems viable for real-world usage and well-performing, reliable, and tailored personal performance.time. Enhancing classification accuracy, improving adaptability, and ensuring long-term usability are essential goals for advancing the practical application of sEMG-PRS, ultimately enabling these systems to meet the demands of real-world usage and deliver reliable, personalized performance.

B. Dataset:

Importing the dataset

To prepare our image dataset for deep learning, we used the powerful tool Keras, allowing data augmentation and preprocessing. We applied such techniques as resizing, rescaling, rotation, zooming, and horizontal flipping, which artificially increase the size of the dataset and add robustness and generalizability to our model.

We imported the dataset of the images directly and read it from a particular folder. The training, validation, and testing data generators were configured separately. The parameters for target image size, batch size, and class mode were specified:

With the dataset pipes well in place, we trained a custom-designed convolutional neural network. The network architecture consisted of convolutional layers used to extract features, pooling layers further reducing the dimensionality, and fully connected layers for classification. By effectively leveraging both augmented data as well as carefully designed architectures, the network was reasonably and finely tuned within the preprocessed dataset to achieve optimal performance.

Training data set:

To train our dataset, we utilize a classifier in conjunction with the fit generator function. This involves defining the number of training steps per epoch, specifying the total number of epochs, and providing validation data along with validation steps. By iterating through these configurations, the model is exposed to a comprehensive dataset, ensuring better generalization and performance. These parameters allow for an efficient and structured training process, accommodating larger datasets and augmentations. Using this approach, the model progressively improves its accuracy and robustness across both training and validation phases.

C. System Architecture

The system architecture for the proposed knee osteoarthritis classification using an ensemble of CNNs consists of a number of crucial layers. As explained in Fig 1, In the Input Layer, images of knee radiographs obtained from medical datasets are preprocessed to ensure homogeneity in size, orientation and resoltuon

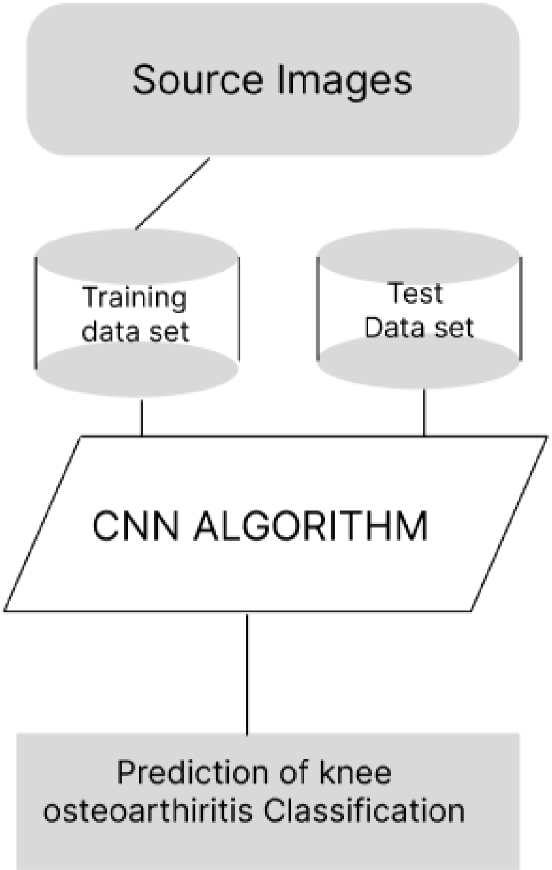

Figure 1: Architectural Workflow

Data augmentation techniques, such as rotation, scaling, and flipping, are applied to increase the variety of images, thereby improving the generalization capability of the model. CNN Ensemble Layer The CNN Ensemble Layer makes use of a variety of different CNN models with different architectures to be able to obtain features across a wide range. For instance, Model 1 captures higher level patterns using certain sizes of kernels and associated pooling strategies, Model 2 uses another configuration of convolution and pooling layers to obtain complementary features, and Model 3 further diversifies feature extraction using unique configurations of layers. Each CNN model comprises convolutional layers

In the Ensemble Layer, as shown in fig.2, outputs of each CNN model are combined through methods such as majority voting or weighted averaging, providing an aggregation of predictions into a single, far more accurate classification output. Ensembles increase robustness and minimise any weak points of a single model. The Classification Layer takes this aggregated output to supply a final prediction on the existence and severity of knee osteoarthritis, which helps in diagnosis. In making the Django Web Application Interface accessible to users, it allows users' health professionals to upload images, view predictions, and diagnostic insights. Moving forward, the Deployment Layer.

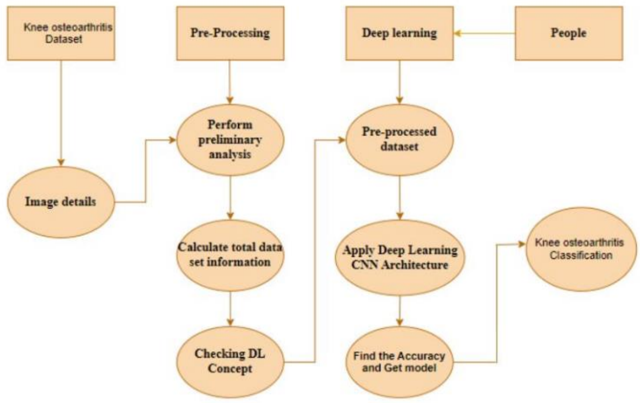

Figure 2: System architecture

As pointed out in Fig.3, ensures that the system is scalable using distributed computing by TensorFlow, therefore well-suited for large-sized datasets in different clinical environments. Secure data storage further enhances the integrity of the data and allows for further in-depth analysis if needed. Added advanced user authentication and data encryption mechanisms ensure meeting healthcare data privacy standards, thus improving the reliability and trustworthiness of the system. This architecture provides a complete, robust, and user-friendly system optimized for real-world clinical use. The CNN ensemble model aims through these advancements not only to set a new benchmark in knee OA classification but also contribute to the broader adoption of AI-driven tools in healthcare.

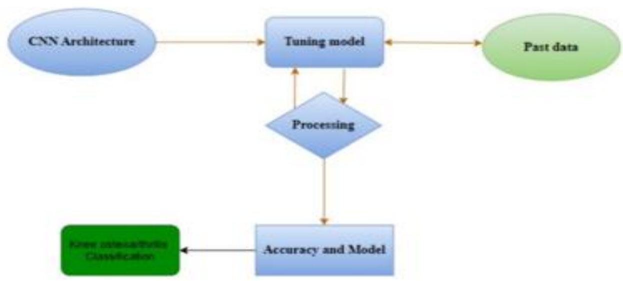

Figure 3: Model Architecture

IV. RESULTS AND DISCUSSION

A. *Experimental* Setup:

This project was developed on a Windows 11 system with Ryzen 5 3500U processor, 8GB RAM, and 512 GB SSD. The application is built using Python 3.11 and Jupyter Notebook from Anaconda 3. The model creation and selection were carried out with the help of the scikit-learn library, while the framework of building the web application was provided through APIs and HTML/CSS with Flask.

B. Observations:

Analyzing the dataset yields several key points. For instance, for numerical features, the patients with knee osteoarthritis had an average age of about 55, and the average BMI of the entire cohort was about 29.4, which indicates a very high proportion of overweighed individuals. In the dataset, there are records of 101 different patients and 10 key biomarkers, among which joint stiffness appeared to be the most common symptom.

Analyzing the progression patterns, high BMI and age were highly associated with the severity of osteoarthritis in which those patients classified as overweight had the faster progression rate. In addition, cartilage degeneration was seen across all ages but proceeded at a faster rate in those who had joint injury histories. Between 2010 and 2020, a drastic increase in the prescription rates of NSAIDs for pain control was noted, as can be seen in Figure 3.

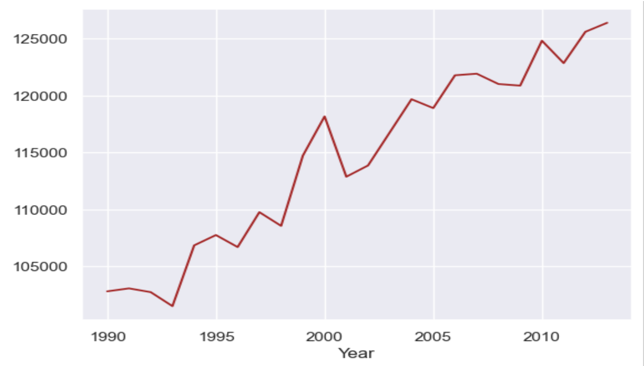

Figure 4: Increasing Cases of knee arthritis

As shown in Fig.4, In addition to knee osteoarthritis, even more thorough examinations of data showed a drastic increase in the number of cases reported in India while the list of reporting countries includes France, Germany, and Japan. There is a high rate of prescription with anti-inflammatory medication ranging between 1,00,000 and 1,30,000 in doses per year.

Table 1: Comparison table for different models

According to experimental findings, Random Forest stacked ensemble model of Gradient Boosting, Bagging and RNN got the lowest RMSE of 9274.09 and best R-squared value of 0.989, making it most accurate to the given dataset.".

It can be also seen that good performance values have been attained from the base models such as Random Forest, Bagging Regressor & Gradient Booster individually, hence a stacked ensemble of these models has been taken along with a deep learning RNN model as meta-learner.

It makes the architecture much more extensible and versatile with ostheo arthiritis prediction performed minimum variance and maximum adaptability, which does not exist in using classic learning models.

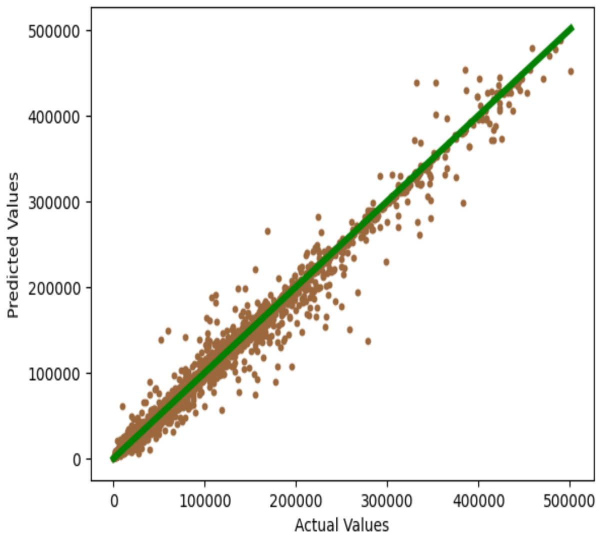

Figure 5: Scatter-Plot for CNN model

V. CONCLUSION AND FUTURE WORK

The proposed CNN ensemble for the classification of knee OA is designed with the idea of overcoming the limitations seen in traditional methods of diagnoses, such as variability in interpretation and reliance on subjective assessments. The strength of multiple CNN architectures is strengthened by combining diverse feature extraction capabilities that serve to boost diagnostic accuracy, consistency, and efficiency. The ensemble model may demonstrate great potential for enhancing early detection and ultimately contributing to better treatment plans and, therefore, better health outcomes and quality of life by providing a robust and scalable framework for OA detection.

Several new research directions will be aimed at improving and expanding the utility of the model. Firstly, infusion of additional data sources - in particular, information on patient history, laboratory work, and general data on patients, like age, gender, and other lifestyle preferences - will be made a priority. Such improvements should enrich the model's predictive capabilities, allowing a more comprehensive analysis of patient-specific risk factors and disease progression patterns to be conducted.

Furthermore, work will be aimed at demonstrating and testing the model in real-world clinical environments. This will include stress testing to ensure compliance with healthcare authorities' standards on the use of AI for diagnosis, as well as other regulatory standards. As part of the integration process, adaptation of the model to existent healthcare workflows will be a crucial need in ensuring smooth interoperability with electronic health record systems and other medical infrastructure. Further research will involve refining the ensemble model with more sources for enhancement in predicting capabilities, such as clinical parameters and patient demographics. In addition, the model will be implemented in various clinical settings in an attempt to be implemented towards real-world practice. It will then meet the assessment of applicable regulatory standards and be integrated into the prevailing healthcare workflow. Longitudinal studies will also be followed in further investigating the model's performance and impact on the updating or refreshing patient management strategies in knee osteoarthritis.

References

1. *Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M.A., & Khan, N. (2021). A comprehensive review of kneeostheo arthritis prediction using machine learning approaches with special emphasis on palm oil ostheo arthritis prediction. IEEE Access, 9, 63406-63439.*
2. *Keerthana, M., Meghana, K. J. M., Pravallika, S., &Kavitha, M. (2021, February). An ensemble algorithm for kneeostheo arthritis prediction. In 2021 Third International Conference on Intelligent Communication Technologies andVirtual Mobile Networks (ICICV) (pp. 963-970). IEEE.*
3. *P. Kumar, S. Senthil Pandi, T. Kumaragurubaran and V. Rahul Chiranjeevi, "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 159-163, doi: 10.1109/AUTOCOM60220.2024.10486087.*
4. *Prasad, N. R., Patel, N. R., & Danodia, A. (2021). Kneeostheo arthritis prediction in cotton for regional level using random forest approach. Spatial Information Research, 29, 195-206.*
5. *Shidnal, S., Latte, M. V., & Kapoor, A. (2021). Kneeostheo arthritis prediction: two-tiered machine learning model approach. International Journal of Information Technology, 13, 1983-1991.*
6. *Raja, S. P., Sawicka, B., Stamenkovic, Z., & Mariammal, G. (2022). Kneeprediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. IEEE Access, 10,23625-23641.*
7. *Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops ostheo arthritis prediction based on machine learning models: Case of West African countries. Smart Agricultural Technology, 2, 100049.*
8. *A. K, S. Y. Prathima, S. Ps, V. V and T. M, "Toxicity Detection in Soap using Deep Learning," 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICSES60034.2023.10465402.*
9. *Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021).Machine learning for large-scale kneeostheo arthritis forecasting. Agricultural Systems, 187, 103016.*
10. *Shidnal, S., Latte, M. V., & Kapoor, A. (2021). Kneeostheo arthritis prediction: two-tiered machine learning model approach. International Journal of Information Technology, 13, 1983-1991.*
11. *Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for kneeostheo arthritis forecasting. Computers and Electronics in Medical, 206, 107663.*
12. *Gnanavel, S., Duraimurugan, N., Jaeyalakshmi, M. (2024). Smart Surveillance System and Prediction of Abnormal Activity in ATM Using Deep Learning. In: Namasudra, S., Trivedi, M.C., Crespo, R.G., Lorenz, P. (eds) Data Science and Network Engineering. ICDSNE 2023. Lecture Notes in Networks and Systems, vol 791. Springer, Singapore. https://doi.org/10.1007/978-981-99-6755-1_11.*
13. *Zhang, Q., Zhao, X., Han, Y., Yang, F., Pan, S., Liu,Z., ... & Zhao, C. (2023). Maize ostheo arthritis prediction using federated random forest. Computers and Electronics in Medical, 210, 107930.*
14. *Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Medical Ostheo arthritiss Based on Machine Learning Using Regression and Deep Learning. IEEE Access.*
15. *Shafi, U., Mumtaz, R., Anwar, Z., Ajmal, M. M., Khan, M. A., Mahmood, Z., ... & Jhanzab, H. M. (2023). Tackling food insecurity using remote sensing and machine learning based kneeostheo arthritis prediction. IEEE Access.*
16. *E Ringdahl, S Pandit - American family physician, 2011 - aafp.org*
17. *K Rönn, N Reischl, E Gautier, M Jacobi - Arthritis, 2011 - Wiley Online Library*
18. *JC Mora, R Przkora, Y Cruz-Almeida - Journal of pain research, 2018 - Taylor & Francis*
19. *Adler-Milstein J, DesRoches CM, Jha AK. Health information exchange among US hospitals. American Journal of Managed Care. 2011;17(11):761–768. [PubMed]*
20. *Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, Edgman-Levitan S, Cleary PD. Examining the role of patient experience surveys in measuring health care quality. Medical Care Research and Review. 2014;71(5):522–554. [PMC free article] [PubMed]*
21. *Agwunobi J, London PA. Removing costs from the health care supply chain: Lessons from mass retail. Health Affairs. 2009;28(5):1336–1342. [PubMed]*
22. *.Bell RM, Koren Y. Scalable collaborative filtering with jointly derived neighborhood interpolation weights; Paper presented at Seventh IEEE International Conference on Data Mining; Omaha, NE. 2007*
23. *Bleustein C, Rothschild DB, Valen A, Valatis E, Schweitzer L, Jones R. Wait times, patient satisfaction scores, and the perception of care. American Journal of Managed Care. 2014;20(5):393–400.*
24. *Gabow PA, Goodman PL. The Lean prescription: Powerful medicine for our ailing healthcare system. New York: Productivity Press; 2014.*
25. *Hall R. Patient flow: Reducing delay in healthcare delivery. New York: Springer Science & Business Media; 2013.*
26. *HHS (Department of Health and Human Services). HITECH Act enforcement interim final rule. 2015. [May 1, 2015]. http://www.hhs.gov/ocr/privacy/hipaa/administrative/enforcementrule/hitechenforcementifr.html.*
27. *Howley MJ, Chou EY, Hansen N, Dalrymple PW. The long-term financial impact of electronic health record implementation. Journal of the American Medical Informatics Association. 2015;22(2):443–452. [PubMed]*