International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.039

Unlocking the Potential of Machine Learning in Osteoporosis Detection: A Comparative Study of Multilayer Perceptron, Convolutional Neural Network and Dropout Models

ShivaramanRamakrishnan
Dept. of Data Science and Business System
SRM Institute of Science and Technology
Chennai, TamilNadu
sr7329@srmist.edu.in

Pavan Nikhil Yeturu
Dept. of Data Science and Business System
SRM Institute of Science and Technology
Chennai, TamilNadu
py7603@srmist.edu.in

Dr Paul T Sheeba
Dept. of Data Science and Business System
SRM Institute of Science and Technology
Chennai, TamilNadu
pault@srmist.edu.in

***Abstract* — Osteoporosis is a chronic condition characterized by low bone density and an increased risk of fractures. It is a common problem among older adults and can significantly impact the quality of life and healthcare costs. Early diagnosis of osteoporosis is crucial for the implementation of effective treatment and prevention strategies. However, current methods for detecting osteoporosis, such as Dual-energy X-ray Absorptiometry (DXA), are not always reliable and can be inconvenient for patients. In this article, we explore the application of machine learning methods for osteoporosis detection and have evaluated it using a few models, like the Multilayer Perceptron, Convolutional Neural Network A - 3 Conv. layers, (16 32 64), Convolutional Neural Network B – 4Conv. layers, (32 64 128 128) and Convolutional Neural Network+ Dropout (over Dense Layer only) to see which one might give us the best result. We used a dataset of bone scans compiled and categorized by Mendeley data and used this data to train the models. Our results demonstrate that these models can effectively identify individuals with osteoporosis with some accuracy and sensitivity, some more than others. Overall, our study shows that the best-trained model found during cross- validation has an accuracy of 89.34%.**

Keywords—- Osteoporosis, Machine Learning, Deep Learning, Detection, DXA Scans, Bone Density, CNN, Multilayer Perceptron, classifiers

I. INTRODUCTION

Bones support our bodies and give us mobility. They preserve the health of our heart, brain, and other internal organs. Bone is a growing, living tissue. It is mostly composed of collagen, a protein that imparts strength and hardness and calcium, a mineral that offers a soft structure. A calcium deficiency can lead to the bones becoming weak and losing most of their power and rigidity. This leads to them breaking up most of the time, even in minor disturbances. This condition is called osteoporosis. Osteoporosis can cause spinal or hip breaks that may provoke budgetary weight and high depressing ness. As needs are, there is a prerequisite for the early examination of osteoporosis and expecting the proximity of the break. According to the World Health Organization, musculoskeletal problems account for up to one in three cases of disability globally and impact people of all ages. These ailments include more than 150 diagnoses, ranging from straightforward fractures that recover quite fast to chronic illnesses like osteoarthritis. Such musculoskeletal problemswill become more prevalent and their accompanying socioeconomic consequences will rise as the world's population ages. Over the next few decades, the significance of accurately detecting these conditions, frequently done by Bone X-Ray scans evaluated by skilled radiologists, will only increase. An automated method for swiftly and affordably identifying bone abnormalities from X-Rays would be extremely helpful given that there are so few qualified radiologists in the world, with only a handful an automated method for efficiently and affordably identifying bone abnormalities from X-Rays would be extremely helpful given that there are so few qualified radiologists in the world, with only a handful of them being found in underdeveloped nations.

Our main objective is to build a system that can predict from the input if said person has osteoporosis, i.e., his bones are normal or not, and for that, we will train our model using a dataset and effectively train the model to get better at predicting.

This machine learning project aims to experiment and provide different solutions employing various neural network architectures in a binary classification task. In particular, the dataset at hand is composed of 614 images representing normal (359) and osteoporosis (255), a standard image classification dataset for composing and improving neural network models. All the different phases of the work are developed in python with the TensorFlow machine learning library via Colab Notebook.

II. LITERATURE REVIEW.

A recent study conducted on detecting liver disease lesions used a very highly sought-after and upcoming CNN architecture, DenseNet which was trained with around 10000 real-time samples of liver Xrays. The resulting model had an accuracy of 98.34, higher than previous works due to DenseNet‘s unique dropout layer [1].

A couple of researchers from McMaster University in Canada along with a few researchers from other universities conducted a study by training a CNN to identify vertebral compression fractures but with a twist. Instead of using a pre-labelled dataset, they used active learning to decrease

DOI: rp-9788770040723.039

class imbalance and produce an effective image classifier. This reduced the cost and time required to train an ML model [2]. CNN was also used todetect lung nodule candidates by training the model with the LUNA16 dataset. This turned out to be a huge success and the results surpassed the previous SOTA approach. The model also was insensitive to the input sizes of the image making it highly useable across a wide range [3]. To make skin cancer detection easier and more accurate, work was done on combining well-known deep-learning models to extract features and then use those to train support vector classifiers. The resulting model was able to give an accurate prediction of 83.83% for melanoma classification and 97.55% for another classification [4]. Another Skin cancer detection proposal used a CNN on a dataset consisting of a whopping 129,450 images to effectively compete against 21 board-certified dermatologists to classify the images into two critical binary classes. The CNN performed as good as the doctors [5]. Researchers in china concluded that the deep learning model, DCNN, could not replace DEXA for BMD (Bone Mass Density) screening, however, it might be employed if a DEXA has not been performed but a lumbar spine X-ray is easily accessible [6]. Another proposal we came across tested a Multilayer Perceptron and Naive Bayes on 33 scans. The results showed that the Multilayer Perceptron outperforms the Naive Bayer classifier in every way [7]. A test was conducted between SVG and an EBP-NN (Error Back Propagation Neural Network) to see which was able to classify fractured and non-fractured bones accurately using a dataset which was preprocessed with wavelet transform to remove noise. The SVG outperformed the EBP-NN by 2% [8]. There was a test between an FEA model and a CNN model to test bone anonymities. Both were put to the test using a large database of artificially produced cancellous bone anatomy. The execution time difference between the FEA model and CNN was around 1000 times, from 32.1 seconds to 0.03 seconds [9]. This is one of the studies that helped us go with CNN architecture. Twenty machine learning techniques were evaluated based on their popularity and frequency in biomedical engineering challenges to divide subjects into two classes (osteoporosis and non-osteoporosis). The well-known 10-fold cross-validation method has been used to evaluate all classifiers, and the results were presented analytically. Their research showed that "age" and "weight" were rated as the most significant diagnostic criteria initially generating the feature set by a feature selection approach. It was evident that eliminating the "sex" diagnostic component did not affect the majority of techniques' efficacy [10]. Cruz et al. methodically gathered and condensed the main approaches used to categorise risk categories for osteoporosis, highlighting their issues and patterns. In conjunction with earlier studies like QUS and DEXA, methods that used AI principles for categorising risk groups were emphasised, concluding that developing a model utilising AI to forecast risk groups has frequently proven to be very beneficial for the patients in their treatment. A relatively new and non-invasive technique for determining bone mineral state at the peripheral skeleton is quantitative ultrasonography (QUS). In addition to bone density, QUS methods offer some structural data that may be crucial in assessing the risk of fracture [11]. The use of fuzzy neural networks (FNN) to identify postmenopausal women with osteoporosis was suggested in another proposal that we came across. This study used 100 postmenopausal women's dental panoramic radiographs from visits to their clinic for BMD evaluations at the lumbar spine and femoral neck. The results indicate that postmenopausal women with osteoporosis can be identified in the dental clinic using a combination of cortical width and shape by employing FNN. Fuzzy neuralnetworks combine the advantages of fuzzy systems and neural networks, enabling them to consider various characteristics and variables related to a condition and produce more accurate and dependable forecasts. Using the new FNN-based system, dentists can effectively identify postmenopausal women and then refer them for BMD testing to obtain an accurate t score and continue with their testing. [12]. A system to automatically detect and localisetumours as small as 100 x 100 pixels in gigapixel microscopy images with a resolution of 100,000 x 100,000 pixels has been developed in studies on employing CNNs for disease diagnosis, such as this one. On the tough lesion-level cancer identification challenge, their approach, which makes use of a convolutional neural network (CNN) architecture, achieves state-of-the-art results when trained with the Camelyon16 dataset. According to their reported results, CNN detects 92.4% of tumours compared to 82.7% for the previous best-automated method. [13]. A proposal conducted by researchers in Saudi Arabia led them to produce a model using Mask-RCNN that is trained to perform bone age assessment and classify them. Without changing the program's structure, they applied innovative methods like the whale optimization algorithm to handle various optimization challenges for real-world applications. The resulting model had a maximum accuracy of 99.2% [14]. A study was done on assessing how well neural networks have helped various healthcare operatives were documented in a survey, where it is noted that CNN, especially in the field of ML has had the most success in medical image classification among other ML-based solutions [15].

III. Data Pre-Processing

The first step is to load the dataset into the Jupiter notebook to start working on it. Assuming the images are stored locally, this is achieved via the os python library. A crucial first step for all types of machine learning applications is to check the dataset thoroughly and apply some kind of pre-processing even the best model cannot learn much from a poor, noisy and inconsistent dataset. Given the project's experimental nature, all images present in the original dataset are being considered, without any kind of manual removal of noisy samples. Anyhow, the images have different sizes and are stored in jpg format.

DOI: rp-9788770040723.039

Hence, in this case, pre-processing consisted of a resize of all images and a colour conversion, resulting in 250x250 grayscale images. These procedures aim to align the dataset, speed up the training and reduce the memory requirements without affecting the final results too much. This is achieved via a single custom function adopting the computer vision library Open-CV. Not much else applies to images in terms of data pre-processing, however, data normalisation can be employed. It consists of adjusting feature values to have them on a similar scale, this is proven to increase the performance and the training stability of the model, in particular of neural networks. In this case, grey scale pixel values ranging from 0 to 255 are normalized in the 0-1 range.

After this phase, images are stored in an array with shapes (614, 250, 250) and labels are represented in a separate array with shapes (614, 2). Labels are represented via one-hot encoding since classes are not related in any way and distance has no meaning.

Using 5-fold cross-validation, the risk estimate for each model is calculated. This enables us to train the model on several dataset subsets, produce five distinct predictors using a particular network architecture, and assess the overall model performance by taking into account the average loss and accuracy across all predictors. This is accomplished using a custom function that folds the images and their labels simultaneously shuffles the images and their labels, trains the model over four folds, and then tests it over the final fold five times. Naturally, the model weights are reset at the beginning of each training phase, thus five distinct predictors are generated. The function then returns the average loss and accuracy of the predictors as an overall evaluation of the learning algorithm with that fixed structure and hyper parameters.

IV. MODEL SET-UP

Each model is built in Tensorflow. Keras uses the same training loss function and optimizer. Since the approach only attempts to optimise the loss function, it is essential for the model's overall performance. A standard loss function for the binary classification task is employed in this instance, which is binary cross-entropy. The optimizer, which affects the network's updated weights, is also crucial. Adam, an expanded variation of stochastic gradient descent that has become the industry standard for deep learning and computer vision tasks, is employed in this situation.

The batch size and epoch count are additional important training factors. The batch size determines the number of samples the model runs through before updating its internal parameters. An epoch is a full pass over the entire training set, and the number of epochs defines the number of times the algorithm goes through it during the training procedure and so we set the batch size to 64 and the no of epochs to 15.

The accuracy and the zero-one loss are additional parameters that are offered when evaluating each model. The accuracy gives back the total number of accurate predictions across all samples taken into account, typically given as a percentage. The custom metric known as the zero-one loss merely counts the number of errors over the overall sample count taken into account.

The architectures we chose are as follows:

*Multilayer perceptron:*The multilayer perceptron is the first neural network architecture to be tested. It has six layers and is a feed-forward neural network. The first layer is a flattening layer, which converts the 100x100 2D input array into a 10000 by 10000 1D array to enable computations for the subsequent levels. Basic layers with dense connections make up the following four layers; the first has 100 neurons and the second, has 64. The Relu function, which is the most frequently utilised for hidden layers since it prevents the vanishing gradient problem during training, serves as the activation function for these layers. Only two neurons make up the output layer, representing the two distinct predictions' outcomes. The sigmoid activation function is utilised here, which outputs two values between zero and one summing upto one, representing the likelihood of either normal or osteoporosis.

The results were as follows:

MLP Metrics - 5-fold cross-validation estimate:

0-1 LOSS: 16.6 (over batch size = 64).

BCE LOSS: 0.5510249614715577

ACCURACY: 0.7221181392669678

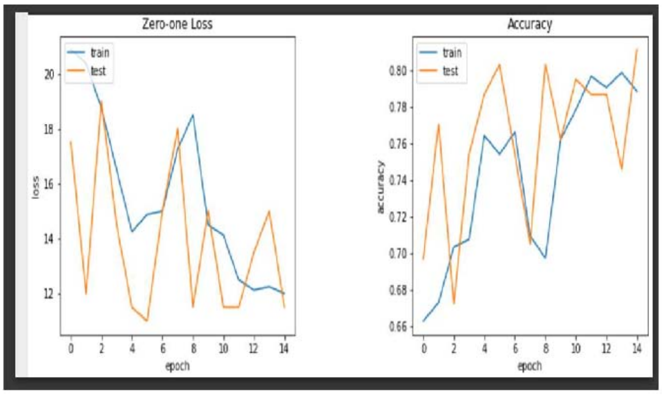

Fig 1. MLP Results

*Convolutional neural network A - 3 Conv. layers (16 32 64):*A convolutional layer followed by a max-pooling layer makes up the building block of a convolutional neural network; the whole network is made up of many blocks of this type. The number of applied filters, size of the kernel, and padding especially to prevent the shape of the picture from being altered during convolution must all be specified when building a convolutional layer. Three blocks with increasing numbers of 3x3 filters (16, 32, 64) were utilised for this model. The output must next pass through a flattening layer to become less dimensional before being fed to a dense layer made up of 512 neurons. Again, there are only two neurons with a sigmoid activation function in the output layer.

DOI: rp-9788770040723.039

The results were as follows:

CNN_A Metrics - 5-fold cross-validation estimate:

0-1 LOSS: 13.2 (over batch size = 64).

BCE LOSS: 0.5615522921085357

ACCURACY: 0.7696174907684326

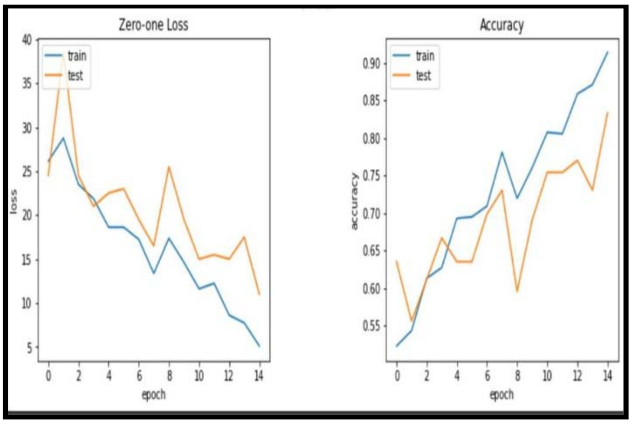

Fig 2. CNN (A) Results

Convolutional Neural Network B - 4 Conv. layers, (32 64 128 128): A second CNN can be defined to do additional research on this powerful architecture. This CNN is made up of four blocks that generate 32, 64, 128, and 128 filters, while the rest of the network is left unaltered. This ought to give the network additional strength and enable it to gather more information.

The results were as follows:

CNN_B Metrics - 5-fold cross-validation estimate:

0-1 LOSS: 13.1 (over batch size = 64).

BCE LOSS: 0.5430635213851929

ACCURACY: 0.7896695256233215

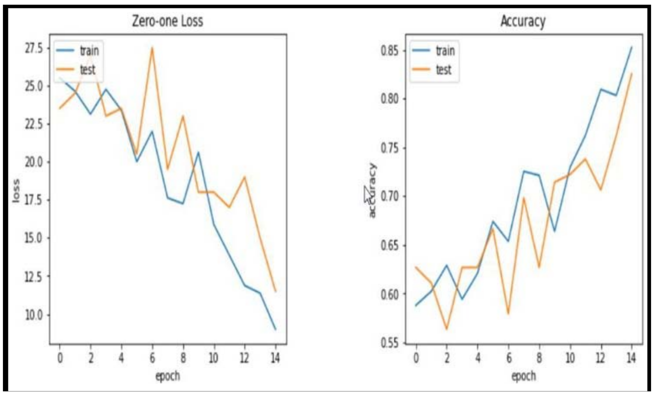

Fig 3. CNN (B) Results

Convolutional neural network + Dense dropout: Here we took the previous B model and added a dropout layer right before the dense layer and the dropout rate was set to 0.2.

The results were as follows:

CNN_dropout_dense Metrics - 5-fold cross-validation estimate:

0-1 LOSS: 11.0 (over batch size = 64).

BCE LOSS: 0.514670866727829

ACCURACY: 0.7984426307678223

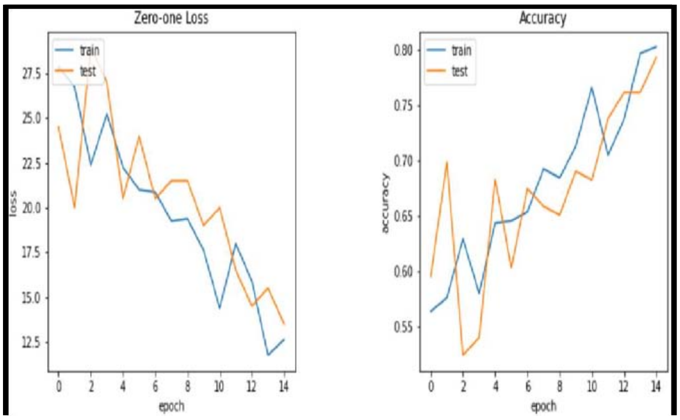

Fig 4. CNN Dropout over Dense layer, rate = 0.2 - Results

All the above graphs refer to a single step of the k-fold CV, graphs from other steps show a similar trend.

These results show that overfitting has been lessened but not completely avoided when compared to those CNN B. With dropout, this occurs around the 9th epoch as opposed to CNN B, when test loss started improving around the 7th epoch.

V. CONCLUSION

In the end, we performed a K-fold Cross Validation to pick the best-trained model and give out the best accuracy encountered. The result was as follows:

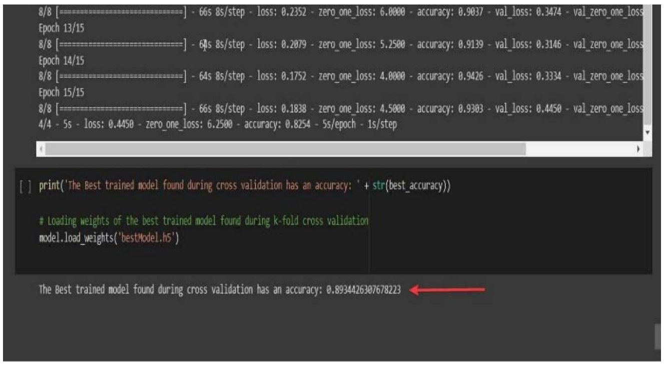

Fig 5. Accuracy of the model

The purpose of this research was to explore various neural network topologies and learn how to conduct tests and assessments using k-fold cross-validation that was statistically sound. Having stated that, the most recent models nearly attained a 90% accuracy. There is still room for improvement but our main goal for the next step of this research is to look at CNN models with a dropout layer

DOI: rp-9788770040723.039

and some particular architecture like DenseNet to take this to the next level.

REFERENCES

[1] N Nanda Prakash, V Rajesh, Syed Inthiyaz, Sandeep D P, SkHasaneAhammad, A DenseNet CNN-based Liver Lesion Prediction and Classification for Future Medical Diagnosis, Scientific African,e01629, ISSN 2468-2276, 2023.

[2] Barret A. Monchka, John T. Schousboe, Michael J. Davidson, Douglas Kimelman, Didier Hans, ParminderRaina, William D. Leslie, Development of a manufacturer-independent convolutional neural network for the automated identification of vertebral compression fractures in vertebral fracture assessment images using active learning, Bone, vol. 161, p. 116427, ISSN 8756-3282,2022.

[3] W. Zuo, F. Zhou, Z. Li and L. Wang, "Multi-Resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection," in IEEE Access, vol. 7, pp. 32510-32521, 2019, doi: 10.1109/ACCESS.2019.2903587.

[4] A. Mahbod, G. Schaefer, C. Wang, R. Ecker and I. Ellinge, "Skin Lesion Classification Using Hybrid Deep Neural Networks," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, pp. 1229-1233, 2019,doi: 10.1109/ICASSP.2019.8683352.

[5] A.Esteva,B.Kuprel, and R. Novoa, et al. “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115–118, 2017,https://doi.org/10.1038/nature21056.

[6] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022). Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks. Journal of Sensor and Actuator Networks, 11(4), 82.

[7] RinishaBagaria, SulochanaWadhwani, Arun Kumar Wadhwani, “Bone fractures detection using support vector machine and error backpropagation neural network,”Optik, vol. 247, p. 168021, ISSN 0030-4026, 2021.

[8] EnnyItjeSela, Reza Pulungan, Osteoporosis identification based on the validated trabecular area on digital dental radiographic images, in: 4th International Conference on Computer Science and Computational Intelligence 2019.

[9] C.F. Cius d̦ el , A. Vizitiu, F. Moldoveanu, C. Suciu, L.M. Itu, Towards deep learning based estimation of fracture risk in osteoporosis patients, in: 2017 40th International Conference on Telecommunications and Signal Processing (TSP), IEEE, pp. 659–665, 2017.

[10] TheodorosIliou, Christos-NikolaosAnagnostopoulos, George Anastassopoulos, “Osteoporosis detection using machine learning techniques and feature selection,” Int. J. Artif. Intell. Tools, vol. 23, no. 05, p. 1450014, 2014.

[11] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications. Computers and Electrical Engineering, 102, 108130.

[12] Arifin AZ, Asano A, Taguchi A, Nakamoto T, Ohtsuka M, Tsuda M, et al., “Developing computer-aided osteoporosis diagnosis system using fuzzy neural network,” J AdvComputIntellIntell Inform, vol. 11, pp. 1049–58, 2007.

[13] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E. Dahl, TimoKohlberger, Aleksey Boyko, SubhashiniVenugopalan, AlekseiTimofeev, Philip Q. Nelson, Greg S. Corrado, Jason D. Hipp, Lily Peng, Martin C. Stumpe, Detecting Cancer Metastases on Gigapixel Pathology Images:https://arxiv.org/abs/1703.02442.

[14] M. A. Duhayyim, A. A. Malibari, M. Obayya, M. K. Nour, A.S. Salama et al., "Metaheuristic with deep learning enabled biomedical bone age assessment and classification model," Computers, Materials & Continua, vol. 73, no.3, pp. 5473–5489, 2022.

[15] M. W. Nadeem, H. G. Goh, A. Ali, M. Hussain, M. A. Khan, and V. a/p Ponnusamy, “Bone Age Assessment Empowered with Deep Learning: A Survey, Open Research Challenges and Future Directions,” Diagnostics, vol. 10, no. 10, p. 781, Oct. 2020, doi: 10.3390/diagnostics10100781.