Detecting Lower Back Pain Using Stacked Ensemble Approach

[1]Shawni Dutta and [2]Prof. Samir Kumar Bandyopadhyay

[1]Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India.

[2] Academic Advisor, The Bhawanipur Education Society College, Kolkata, India.

Abstract-

Lower Back Pain (LBP) is a disease that needs immediate attention. Person with back pain shall go immediately to doctor for treatment. Injury, excessive works and some medical conditions are result of back pain. Back pain is common to any age of human for different reasons. Due to factors such as previous occupation and degenerative disk disease the chance of developing lower back pain increases for older people. It hampers the working condition of people common reason for seeking medical treatment. The result is absence from work and is unable to normal due to pain. It creates uncomfortable and debilitating situations. Hence, detecting this disease at an early stage will assist the medical field experts to suggest counter measures to the patients. Detection of lower back pain is implemented in this paper by applying ensemble machine learning technique. This paper proposes Stacking ensemble classifier as an automated tool that will predict lower back pain tendency of a patient. Experimental result implies that the proposed method reaches an accuracy of 76.34%, f1-score of 0.76 and MSE of 0.34.

Keywords: Lower back pain, automated tool, ML, ensemble technique, stacked generalization.

Introduction-

The back side of human is composed of a complex structure of muscles, ligaments, tendons, disks, and bones. Components of this structure work together to support the body and enable human to move around. The segments of the spine are cartilage-like pads called disks. Back pain may occur due to the problems with any of these components. In some cases, reasons of back pain are still unknown. Lower back caused due to the bony lumbar spine, discs between the vertebrae, ligaments around the spine and discs, spinal cord and nerves, lower back muscles, abdominal and pelvic internal organs, and the skin around the lumbar area. Upper back pain is caused by disordering of the aorta, tumors in the chest and spine [1].

Artificial intelligence (AI) is a field of engineering and it has potential to enhance healthcare through decision making facility of patient engagement. Machine learning (ML) [2] can be used to automate decision making and make predictions based upon patient data. The

causes of LBP are incompletely understood and unresolved medical problems that continue to have a high prevalence. It is accepted that pain is a complex phenomenon. Computational science can use complex clinical and experimental data to better understand the complexity of pain. ML is referred to as a set of methods to detect patterns pain through dataset data and then it can use the uncovered patterns to predict or classify future data, to observe structures such as subgroups in the data, or to extract information from the data suitable to derive new knowledge. Biostatistics, artificial intelligence and machine learning aim at learning from data. So ML methods have the power for analyzing LBP.

ML approaches can explore data sets by reversing the analytical focus of classifier building and pattern detection. It can qualify for data exploration under the assumption that if a biomedical parameter qualifies for inclusion in a classifier, then it is probably important for the addressed pain-related problem. In classic methods knowledge or at least presumptions about the distributions and/or functional dependencies of the data are necessary. ML methods allow for data-driven research approaches. Techniques of feature selection are common to ML for enabling one to identify relevant modulators of pain-related outcomes in data-driven and hypothesis-free explorative research approaches. ML analysis identified pain-related parameters as the most relevant for explaining the persistence of pain in human.

The objective of this study is to detect the feasibility of utilising relevant medical records and conclude the probability of being affected by lower back pain disease. To address the objective of this study, data mining and knowledge discovery approches are explored that will extract the hidden patterns in medical database [3]. Machine Learning (ML) algorithms are utilised for this purpose in order to learn from extracted patterns. ML also enable machine to demonstrate itself to adapt behaviours for forthcoming events. Classifier models are supervised ML techniques that accept patients' records as input and identify lower back pain tendency of a patient. This paper attempts to use ensemble based ML technique. Use of ensemble technique is quite popular since it combines multiple models with reduced variance of prediction and optimised generalization error. Stacked generalization method follows ensemble technique which is superior over cross-validation method. This method exploits scheme of winner-takes-all for combining the individual generalizers [4]. This paper implements a couple of predictive models such as Naïve bayes [5], k-Nearest Neighbour [6], Multi-layer Perceptron [7], AdaBoost [8], Gradient Boost [9], Random Forest [10] classifiers as baseline models. Next, top two models are identified based on evaluation of prediction performance. Finally these two classifiers are assembled under a single platform using stacked generalization ensemble technique. For justifying the evaluation process, accuracy [11], f1-score [12] and MSE [12] are the used as performance measure metrics.

Related Works-

LBP is normally common health problem for aged people. However, LBP might be an issue for middle aged people due to excessive pressure in the work place. It is a major problem of

those are using computer for a long time without any movement. Suppose In an office one has a LBP but his/her co-workers cannot feel it at the initial stage. LBP can occur in all age groups and someone can adjust it in daily life but others may not do so [1].

In [13], numerous classifiers such as Support Vector Machine (SVM), logistic regression (LR), bagging ensemble methods are applied for the diagnosis of spinal abnormality. Before implementing these classifier models, data pre-processing step are carried out. In data pre-processing step, uni-variate feature selection as a filter based feature selection, and principal component analysis (PCA) as a feature extraction algorithm are considered. Experimental results conclude that bagging SVM provided suitable classification results.

Researchers suggest that there exists difference in between the questions that need to be answered by medical practitioners relating to the occurrence of the pain and type of treatment [14]. It is mentioned by the doctors to reduce weight and take essential care.

Researchers proposed an algorithm to classify LBP acquiescent to chiropractic treatment and pains due to pathological cause [15]. The method allows the practitioners to avoid making assumptions depending upon the heuristics and pattern recognition until it has been established that the patient is a candidate for chiropractic care. It considers underlying distinction of four clinical patterns which are simple mechanical lower back pain, lower back pain with radiculopathy, serious pathological lower back pain and lower back pain with psychological overlay.

Individuals in present lifestyle have spent extended periods in a seated position at work as well as during leisure time. The prevalence of low back pain (LBP) has increased among office workers in general. Computer knowledgeable employees in Call Centres have recently become the focus of attention in this field as they spend most of time for doing work in a seated position [16]. Their jobs are also recognized for potentially high levels of stress especially when dealing with difficult or aggressive customers. High job-related stress is related to musculoskeletal disorders of the lower back. So in the higher portion of employees in Call Centres, it is reported musculoskeletal symptoms than other professional office users. LBP can disable a person since many other diseases related to it [17]. In turn it may be a major economic burden to society. The identification of risk factors in office environment needs to give high attention and therefore; it is required to device a method to study LBP using related dataset. The main aim of the paper is inclined to this situation in the Office/Call Centres [17].

Spine supports structure of human body that connects different skeleton parts and keeps body upright [18]. In the older age the bones, discs and ligaments in the spine are naturally weaken. It happens to all of humans to some degree as part of the ageing process. It does not mean that everyone will have pain from this.

ML is based on the availability of data, which is used to train the machine to perform the desired tasks. Due to its nature, ML lends itself well to applications in which input data are used to generate an output based on some features of the inputs themselves, for example, to perform image classification. Indeed, a research area which was dramatically advanced by ML in recent years is image processing [19]. Deep neural network achieved for the first time superhuman performance in a famous image classification contest, the Image Net Large Scale imaging data are interpreted and exploited for treatment planning and follow-up [20]. The impact of ML and AI on other basic medical research fields has been less conspicuous so far; nevertheless, numerous novel applications, for example, in motion analysis and mechanical characterization of tissues, are starting to emerge. As testified by the sharp increase in the number of published papers in recent years, AI and ML are more and more being used to investigate issues related to the spine, especially in radiological imaging but also in other fields such as the outcome prediction of treatments. The reported results are either promising or already surpassing the previous state of the art in several applications; for example, ML techniques nowadays allow for an accurate and perfectly repeatable grading of intervertebral disc degeneration on magnetic resonance imaging (MRI) scans. Indeed, the current pace of technical improvements is expected to being further benefits in the next future.

Proposed Methodology-

The framework designed to detect patients with lower back pain trouble proceeds through multi-step paradigms. This paper attempts to use numerous classifier models those are used to associate input variables into target classes while learning from extracted patterns in the training data. Well-known classifiers such as naïve bayes, K-NN, MLP, AdaBoost, Gradient Boost, and Random Forest are implemented as baseline classifiers. Predictive performance obtained from these baseline classifiers are evaluated with respect to some pre-defined metrics employed in selection phase. The aim of selection phase is to provide top two predictive models. This paper attempts to use stacked generalization based ensemble method which combines the predictions from two best models picked up by selection phase. This section describes the detailed explanation of this entire framework and brief summarization is provided in Fig1.

1. Baseline Classifiers-

The *Naive Bayes* classifier [5] is a supervised classification tool that exemplifies the concept of Bayes Theorem [21] of Conditional Probability. This classifier attains very promising result in the following scenario- when the features are independent or features are completely functionally dependent. The decision made by this classifier is quite effective in practice even if its probability estimates are inaccurate. The accuracy of this classifier is not related to feature dependencies rather than it is the amount of information loss of the class due to the independence assumption is needed to predict the accuracy [5].

K-Nearest Neighbours (K-NN) [6] is a supervised ML algorithm that is often known as Memory based classification. During classification process, it identifies objects based on closest proximity of training examples in the feature space. It is known as lazy learners because during training phase it just stores training samples. The classifier considers k number of objects as the nearest object while determining the class. The main challenge of this classification technique relies on choosing the appropriate value of k. While calculating distance among instances, Minkowski distance [6] is considered to be general metric for any data.

Multi-layer perceptron (MLP) [7] can be used as supervised classification tool by incorporating optimized training parameters. MLP follows feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. For a given problem, the number of hidden layers in a multilayer perceptron and the number of nodes in each layer can differ. Fine-tuning the interfering parameters of the network architecture relies on problem-specific domain [7].

Boosting is an efficient technique that is applied on combination of multiple unstable learners in order to improve accuracy of classification. *AdaBoost* is known to be the first boosting technique proposed by Freund and Schapire [8]. This algorithm also belongs to the category of interpolating classifiers which defines algorithmic property of fitting the training data completely without error. This algorithm is a promising example of boosting technique that produces improved output even when the performance of the weak learners is inadequate [8].

Gradient Boost algorithms [9] are boosting techniques those are suitable in fitting new models to provide maximised efficiency while estimating response variable. The objective of this algorithm is to construct new base learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble. This algorithm is highly customizable to any domain which provides freedom in model designing. One of the important issues of this algorithm is identifying and incorporating loss function to this algorithm which is subject to change as a matter of trial and error [9].

Random forest (RF) [10] exemplifies the concept of ensemble learning approach and applies regression technique for classification based problems. This classifier is a combination several tree-like classifiers which are applied on various sub-samples of the dataset and each tree cast its vote to the most appropriate class for the input [10].

Implementation of Baseline Classifiers-

The k-NN classifier receives value of k as 5. The Minkowski distance is used while calculating distances among instances. The MLP classifier is implemented with four hidden layers with sizes 128, 32, 16, 8 respectively. The NB classifier is designed as Multinomial naïve bayes model. The RF classifier is implemented with 800 base estimators. The AdaBoost classifier is

designed using 500 base estimators. The learning rate of this classifier is set to 1.0. The GB classifier is implemented with 500 base estimators, learning rate of 0.9.

2. Selection phase using Evaluating Metrics-

The prediction performance of model needs to be evaluated with respect to some pre-defined metrics. The target of selection phase is to obtain top two classifier models those provide maximised performance with respect to employed metrics.

Accuracy [11] is a metric that ascertains the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough metric for evaluating model's performance since it does not consider wrong predicted cases with different weights. For compensating the above mentioned problem, we consider two more metrics known as, Recall and Precision. *Precision* [12] identifies the ratio of correct positive results over the number of positive results predicted by the classifier. *Recall* [12] denotes the number of correct positive results divided by the number of all relevant samples. *F1-Score* or *F-measure* [12] is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall. *Mean Squared Error* (MSE) [12] is another evaluating metric that measures absolute differences between the prediction and actual observation of the test samples. Classifier models that exhibit maximised accuracy, f1-score value and optimised MSE value signify the best predictive model.

3. Proposed Stacked Ensemble Method-

Stacked generalization ensembles dissimilar machine learning methods, each of them estimates its own generalizing biases based on given training set and then filter out those biases. This model follows usage of two levels of models- level 0 models and level 1 model. Several base models are accommodated in level 0 whereas level 1 model consists of one meta-model. Stacked generalization ensures that meta-model learns from the predictions of level-0 models. Generally introducing meta-model over the level-0 models obtains more accurate prediction over the best level-0 model. The stacking ensemble classifier requires two levels to be present within the model. Use of level-1 model makes stacked generalization method more efficient [4].

In this methodology, selection phase provides two best models which are assembled in these ensemble classifiers. In case of stacked generalization method, these two models are used as level-0 base models and one of them is used as meta-model. The best predictive model is recommended for attaining early prediction of lower back pain identification system.
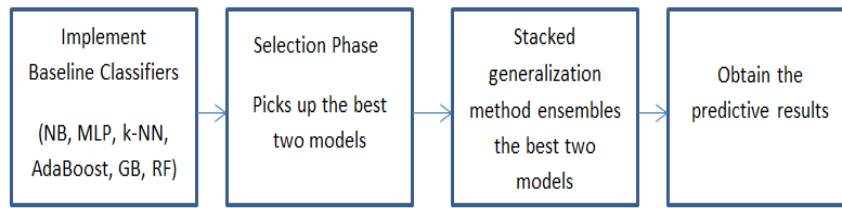
Fig.1: Summary of Implemented framework

Dataset Used-

Lower Back pain symptoms dataset is collected from kaggle and utilized for detecting patients with spine abnormalities. The dataset contains the attributes such as pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, degree of spondylolisthesis, pelvic slope, direct tilt, thoracic slope, cervical tilt, and scoliosis slope. All these data considered as input variables for classifier model. The column 'class_att' present in the dataset signifies normal patients and patients with abnormalities. This attribute is kept as dependent variable during classification task. Fig2 depicts the distribution of target or dependent variable in the dataset. This dataset contains no missing values such as Nan (Not-a-Number) which eliminates the overhead of considering unknown value handling task. The attribute values present in the dataset is scaled down into the range from 0 to 1 using Min-Max-Scaler function. Once the pre-processing is over, the dataset is partitioned into training and testing dataset with a ratio of 7:3. The classifier model accomplishes its learning from extracting patterns from training data and later prediction is obtained for testing dataset.
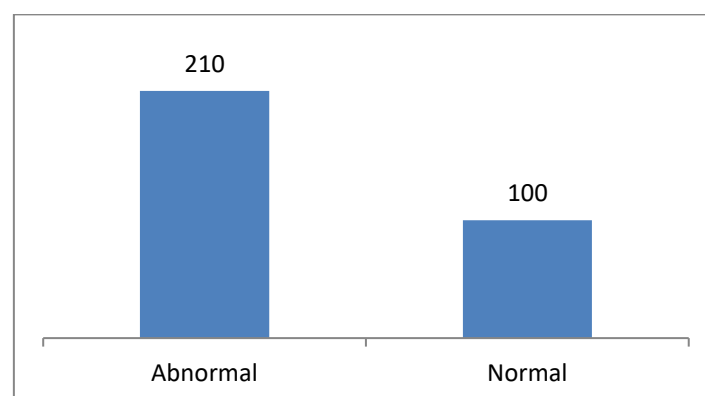

Fig.2: Distribution of Normal and Abnormal Patients in the Dataset

Experimental Results-

The proposed method as well as the baseline classifiers is evaluated with respect to accuracy, f1-score and MSE. The prediction performances of all specified classifier model are shown in Table 1.

Performance Measure Metrics	Accuracy	F1-Score	MSE
Baseline Classifiers			
Naïve Bayes	63.44%	0.63	0.37
MLP	68.82%	0.69	0.31
AdaBoost	69.89%	0.7	0.3
K-NN	70.97%	0.71	0.29
Gradient Boost	72.04%	0.72	0.28
Random Forest	75.27%	0.75	0.25
Proposed Method			
Stacking Ensemble	76.34%	0.76	0.24

Table 1: Comparative Study of Classifier Models

From the aforementioned comparative analysis drawn in Table 1, it is quite evident that the ensemble based Random Forest Classifier provides best result with respect to all the evaluation metrics. Next, Gradient Boost Classifier also provides relatively better result than other baseline classifiers except Random Forest Classifier. These two classifiers are chosen as input for implementing stacked generalization model. In level-0 of proposed model receives Gradient Boost and Random Forest and level-1 receives Random Forest classifier for prediction purpose. Analysis provided in table 1 indicates that Random Forest classifier gives more promising result over Gradient Boost classifier model. This is why it is selected as level-1 or Meta model for obtaining more accurate result. The target of this classifier is to improve the performance by combining the best two classifiers obtained during selection phase. In this case, predictions of Random Forest Classifier and Gradient Boost Classifier are assembled using ensemble methods to get the best out of the performance of prediction tool.

Conclusions-

Early detection of lower back pain shows significant role in the field of health diagnosis systems. For assisting diagnostic system an automated prediction has been carried out using classification algorithms. Several classifiers are utilised while predicting lower back tendency of a patient. Implementations of these classifiers are accompanied by fine-tuning the hyper-parameters. Adjusting these parameters into correct values will support in attaining the best predictive model. Top two classifiers are selected and utilised for designing stacked ensemble method. In conclusion, the proposed ensemble technique attains the lowest error rate along with highest efficiency which might be the best choice of algorithm for this problem and prediction of the considered disease.

References-

[1]   T. Paolucci, C. Attanasi, W. Cecchini, A. Marazzi, S. V. Capobianco, and V. Santilli, "Chronic low back pain and postural rehabilitation exercise: A literature review," *J. Pain Res.*, vol. 12, pp. 95–107, 2019, doi: 10.2147/JPR.S171729.

[2]   R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine Learning An Artifical Intelligence Approach*, vol. 2, no. January 1983. 1983.

[3]   scar Marbn, G. Mariscal, and J. Segovi, "A Data Mining & Knowledge Discovery Process Model," *Data Min. Knowl. Discov. Real Life Appl.*, no. May 2014, 2009, doi: 10.5772/6438.

[4]   D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[5]   I. Rish, "An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier," no. January 2001, pp. 41–46, 2014.

[6]   P. Cunningham and S. J. Delany, "K -Nearest Neighbour Classifiers," *Mult. Classif. Syst.*, no. May, pp. 1–17, 2007, doi: 10.1016/S0031-3203(00)00099-6.

[7]   F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5–6, pp. 183–197, 1991, doi: 10.1016/0925-2312(91)90023-5.

[8]   R. E. Schapire, "Explaining adaboost," *Empir. Inference Festschrift Honor Vladimir N. Vapnik*, pp. 37–52, 2013, doi: 10.1007/978-3-642-41136-6_5.

[9]   A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.

[10]   L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004.

[11]   P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: An overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000, doi: 10.1093/bioinformatics/16.5.412.

[12]   H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.

[13]   M. Raihan-Al-Masud and M. Rubaiyat Hossain Mondal, "Data-driven diagnosis of spinal abnormalities using feature selection and machine learning algorithms," *PLoS One*, vol. 15, no. 2, pp. 1–21, 2020, doi: 10.1371/journal.pone.0228422.

[14]   Minne HW; Leidig G; Wuster C; et al, "developed spine deformity index (SDI) to qaantitate vertebral crush fractmres ia patients with osteoporosts," vol. 38, p. 1693, 1988.

[15]   L. J. Melton, S. H. Kan, M. A. Frye, H. W. Wahner, W. M. O'fallon, and B. L. Riggs, "Epidemiology of vertebral fractures in women," *Am. J. Epidemiol.*, vol. 129, no. 5, pp. 1000–1011, 1989, doi: 10.1093/oxfordjournals.aje.a115204.

[16]   Michalski RS, Carbonell JG, Mitchell TM, "Machine Learning: An Artificial Intelligence Approach". Berlin, Germany: Springer Science & Business Media; 2013.

[17]   Shalev-Shwartz S, Ben-David S, "Understanding Machine Learning: From Theory to Algorithms", Cambridge, UK: Cambridge University Press; 2014.

[18]   Schmidhuber J. "Deep learning in neural networks: An overview.", Neural Netw.2015;61:85-117.

[19]   Van Tulder MW, Cherkin DC, Berman B, et al," Acupuncture for low back pain (Cochrane Review)." The Cochrane Library, 1999.

[20]   Van Tulder MW, Koes BW, Bouter LM, "Conservative treatment of acute andchronic nonspecific low back pain A systematic review of randomized controlled trials of the most common interventions",Spine 1997,22:2128–56.

[21]   D. E. Walters, "Bayes's Theorem and the Analysis of Binomial Random Variables," *Biometrical J.*, vol. 30, no. 7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710.

[23]   sammy123 (2016, August). Lower Back Pain Symptoms Dataset,Version 1.Retrieved June 02,2020 from https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset