Deep Learning Assisted Medical Insurance Data Analytics With Multimedia System

Cheng Zhang[1]*, B. Vinodhini[2], Bala Anand Muthu[3]

[1] School of Law, Jiangnan University, Wuxi 214122 (China)
[2] Assistant Professor, SNS College of Technology, Coimbatore - 641035 (India)
[3] Associate Professor, Department of Computer Science, Adhiyamaan College of Engineering (India)

Received 14 July 2020 | Accepted 18 April 2021 | Published 27 January 2023

Abstract

Big Data presents considerable challenges to deep learning for transforming complex, high-dimensional, and heterogeneous biomedical data into health care data. Various kinds of data are analyzed in recent biomedical research that includes e-health records, medical imaging, text, and IoT sensor data, which are complex, badly labeled, heterogeneous, and usually unstructured. Conventional statistical learning and data mining methods usually require first to extract features to acquire more robust and effective variables from those data. These features help build clustering or prediction models. New useful paradigms are provided by the latest advancements based on deep learning technologies for obtaining end-to-end learning techniques from complex data. The abstractions of data are represented using the multiple layers of deep learning for building computational models. Clinician performance is augmented by the prospective of deep learning models in medical imaging interpretation, and automated segmentation is used to reduce the time for the diagnosis. This work presents a convolution neural network-based deep learning infrastructure that performs medical imaging data analysis in various pipeline stages, including data-loading, data-augmentation, network architectures, loss functions, and evaluation metrics. Our proposed deep learning approach supports both 2D as well as 3D medical image analysis. We evaluate the proposed system's performance using metrics like sensitivity, specificity, accuracy, and precision over the clinical data with and without augmentation.

Keywords

Convolution Neural Network, Deep Learning, Image Regression, Medical Image Analysis, Segmentation.

DOI: 10.9781/ijimai.2023.01.009

I. Introduction

According to recent reports, the present society produces data more quickly than in any other decade, making numerous doors open for various prediction strategies and making it difficult for analysts [1]. Multiple industries become progressively dependent on excellent data quality, and the interest in the sound factual examination of these data is rising in like manner. In the insurance sector, the provision of information has always been believed to be relevant. The insurance provider shall be entitled to the cases arising from this Arrangement in the form of an agreement to provide a client and shall retain reserves to cover any future obligations. For all possibilities taken into account, the insurance premium has to be paid before the real costs are identified. It is referred to as the reversal of the creation cycle. It infers that the exercises of reserving and pricing are firmly interconnected in actuarial practice. From one perspective, statisticians need to decide a reasonable cost for the insurance items they need to pay. Setting the excellent levels charged to the insured's is done in a data-driven way where prediction models are fundamental and essential.

Hazard based estimating is essential in a serious and well-working insurance market. An insurance organization must then protect its solvency and hold money to satisfy outstanding liabilities [2]. Therefore, holding statisticians should foresee, with most extreme exactness, the aggregate sum expected to pay guarantees that the insurance provider has legitimately conceded to cover for. These reserves structure the primary thing on the risk side of the insurance organization's financial record and accordingly have a significant economic effect. This exploration aspires to advance new, precise, and accurate models for the actuarial work field. Non-life (for example, engine, fire, obligation), life, and well-being backup plans are continually faced with the difficulties of quickly expanding facilities for information assortment and data collection, stockpiling, and investigation. Anyway, utilizing their best-in-class approaches for the insurance business won't have the option to plan a satisfactory reaction to these difficulties and associations with the controls of measurements and considerable information investigation.

Besides, the expanded spotlight on inward hazard and the changing administrative rules encourages the importance of improved apparatuses for actuarial modeling of the prediction models. Specifically, the European Solvency II Directive1 forces new necessities to upgrade policyholder security. With these new administrative rules' ongoing presence, the estimation of future incomes and their vulnerability turns out to be progressively significant [3]. Simultaneously, actuarial prediction models need to consent to existing and pending guidelines. Throughout strategies to survive with complex problems, conventional

* Corresponding author.
E-mail address: sunday@jiangnan.edu.cn

machine learning approaches are not adequate. High performance deep learning computing offers the ability to manage massive medical image data for precise and efficient diagnosis. Deep learning helps to pick and extract characteristics and create new ones. It diagnoses the illness and the predictive objective and offers actionable prediction models to effectively support doctors. The Gender Directive2 has denied the utilization of sex as a hazard factor in insurance estimating, and antidiscrimination laws may advance sooner rather than later, further constraining the legally binding opportunity of insurance organizations.

Demonstrating guarantee losses – alternatively called guarantee sizes or severities – is urgent when evaluating insurance items, deciding capital necessities, or overseeing dangers inside money related establishments. For example, different essential circulations, the gamma or lognormal, have been utilized to demonstrate nonnegative losses [4]. These parametric conveyances are not frequently suitable for actuarial information, which might be multimodal or substantial. Besides, while building aggregate hazard models or joining actuarial dangers from numerous business lines, these serious appropriations don't prompt a systematic structure for the relating total loss circulation. While numerical or recreation calculations are accessible, it is considered advantageous to use routine procedures whenever the situation allows. There is continuously a tradeoff between scientific effortlessness from one perspective and practical demonstrating modeling on the other. Fig. 1. shows the essential features involved in modeling insurance data. It offers the basic features involved in insurance data collection, analysis, and modeling the same.

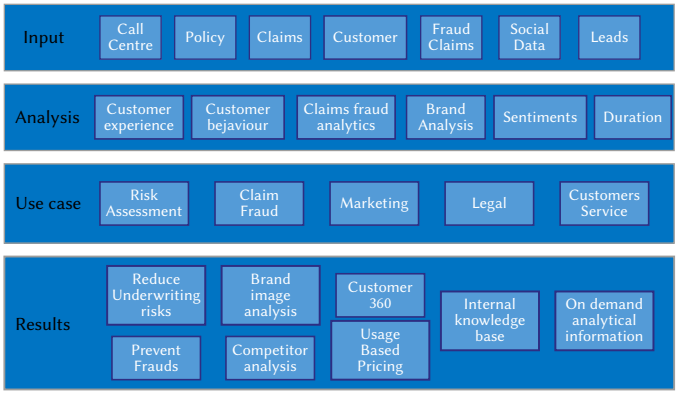

Fig. 1. Insurance Data Modeling.

These days, insurance organizations monitor all kinds of data for every individual case. Rich information sources record, for instance, the event occurrence date, the announcement of a reporting delay, the date and measure of every payment loss, and the settlement date. Many recent forensics advancements [5] monitor the insurance company databases for data theft and attacks over insurance data stored in the cloud [6]. The current strategies for claims holding are intended for totaled information, yet numerous valuable data is lost through this information compression. For considering all the useful information for analytics and establish new relations, deep learning models are considered. In any case, data accessible to play out the prediction work originates from two sources; a group of client characteristics (e.g., client socioeconomics and insurance enlistment data) and the historical backdrop of client emergency clinic affirmation claims. Big data can potentially expose connections, secret trends and other observations by analyzing massive data sets. And data from the learning of machine healthcare led to the discovery of human genomes or medicines for life threatening diseases such as cancer.

The last can be considered as a period of inconsistent occasions. Besides, every hospitalization occasion would itself be communicated as a succession of emergency clinic administrations used by the patient during that specific stay. While it is clear to straightforwardly introduce the principal kind of data as features to a characterization calculation, data mining of time arrangement and point forms, for example, the hospitalization occasions depicted above, is yet an ongoing research territory [7] [8]. It is somewhat testing to remove irrelevant data in a valuable way. The general planning of recorded medical clinic visits for prediction appears to contain unrelated yet significant data that may prompt progressively exact forecasts if it is very well used for insurance data creation.

II. Related Works

Shi et al. [9] designed HFDA, an ensemble Artificial Intelligence way to deal effectively and distinguish the clinical insurance claims implemented in an online clinical insurance claim framework in China. Tooth et al. [10] proposed another client profitability technique for the insurance business by including risk reserve. The proposed scheme can gauge the genuine insurance client commitment viably by thinking about the recorded buying conduct and foreign income. Wang et al. [11] proposed a novel deep learning model for accident coverage extortion discovery that utilizes Latent Dirichlet Allocation (LDA)- based content examination. In the proposed strategy, LDA is first used to detect the content features covering up in the accident cases' content depictions. Deep neural systems are prepared on the information, incorporating the content features and conventional numeric features for distinguishing fake cases.

Koutsomitropoulos et al. [12] developed OWL metaphysics to recognize insurance forms and delineate information volumes gathered in customary information stores. Under thinking, many semantic questions were shown utilizing the vocabularies in the ontology that can simplify examination and derive understood realities from this information. Lin et al. [13] developed a heuristic bootstrap testing approach joined with the hybrid learning calculation for the insurance business data mining. A parallel analysis that utilizes the equal registering ability and memory cache system improved by Spark and used F-Measure and G-intend to assess the calculation's accuracy. The insurance business information from China Life Insurance Company is used to investigate the proposed model efficiency.

Ren et al. [14] proposed a survival forecast model dependent on graph pattern mining. In the first place, every patient's medical coverage information is developed as a Heterogeneous Information Network (HIN). At that point, visit designs are mined from these HINs, and each successive example is viewed as an element called "design highlight." Finally, the survival time is given by an improved random forest, which can consider the edited information from the graph data mining. The investigation is carried out on a genuine medical coverage informational collection to investigate the utilization of factual strategies to make a standard based heuristic motor that works with self-learning Decision Trees. Rayan et al. [15] present an ensemble system that consolidates expertise in the domain and unsupervised learning procedures to recognize false cases. The examination group is implied with a weighted module of extraordinary circumstances posting the most probable fraud cases with proactive and review investigation comments.

Chae et al. [16] inspected the information disclosure attributes and data mining calculations to investigate how they can produce accurate results and give regular data to hypertension prediction using the Korea Medical Insurance Corporation database. In particular, this examination approved the intensity and core ability of data mining calculations by contrasting the presentation of the decision tree, logistic regression, CHIAD (Chi-squared Automatic Interaction Detection), and C5.0 (a variation of C4.5) utilizing the test set of 4588

recipients and the evaluation set of 13,689 recipients. Viveros et al. [17] addressed the viability of two data mining methods in breaking down and recovering unknown personal conduct standards from gigabytes of information gathered in the medical coverage industry. A scene (claims) database for pathology administrations and a general professionals database were utilized for the analysis. Affiliation rules were applied to the scene database; neural segmentation was used to overlay the two databases. The outcomes acquired from this investigation show the potential estimation of data mining in medical coverage data frameworks by distinguishing designs in the pathology administrations and arranging the general professionals into groups mirroring their practices' nature and style. The methodology produced a higher percentage of results that couldn't have been acquired utilizing traditional procedures.

Jiang et al. [18] presented four significant difficulties existing in artificial intelligence in genuine business models. Standard artificial intelligence calculations can commonly be applied to traditional informational indexes, which are ordinarily homogeneous and adjusted. A proficient cost-sensitive parallel learning framework (CPLF) was used to improve insurance tasks with an in-depth learning approach that doesn't require preprocessing. The methodology contains a novel, unified; start to finish a cost-effective neural system that learns genuine heterogeneous information. An explicitly structured cost-delicate grid that consequently produces a powerful model for understanding minority arrangements and the parameters of both the cost-effective lattice and the half breed neural system is, on the other hand, yet mutually upgraded during processing. CPLF-based design for a certifiable insurance knowledge activity framework showed a misrepresentation discovery and arrangement restoration during this framework's investigation. Wang et al. [19] applied a linkage of the Knowledge Discovery in Databases (KDD) procedure to examine the call community information of the NHIA. The practical techniques of handling, determination, data mining, and assessment for two kinds of data mining investigations: information arranging and information affiliation, Moreover, the investigation results and counsel experts in NHIA for the expert assessment about those outcomes and existing medical coverage arrangements showed the establishment of using intelligence paradigms in the medical insurance field. Senthil Murugan and Usha Devi [20] – [21] have proposed a hybrid classification technique for analyzing many data. Additionally, decipher, and reach determinations from these outcomes using information representation are presented for supportability.

Sato et al. [22] built a novel examination system dependent on past investigations that led with clinical scientists and structured a UI that encourages the development of the preparation rationale in a basic and straightforward way. By indicating the execution consequence of commonplace investigations of insurance information, the created system's viability is examined. Umemoto et al. [23] meant recognizing the changes in the prescription patterns and distinguishing its motivation through an investigation of Medical Insurance Claims (MICs), which involve the details of clinical expenses charged to medical safety insurance providers. The methodology is two-crease. Firstly, proposed an inactive variable model that recreates doctors' drug conduct to precisely imitate month to month medicine time arrangement from the MIC information, where medicine interfaces between the illnesses and meds are absent. Secondly, applied a state-space model with intercession factors to deteriorate the month-to-month remedy time arrangement into various parts, including regularity and auxiliary changes. Testing is fair to extract unnecessary information practically. If it is very well used for insurance data production, the general planning of reported medical clinic visits for prediction tends to include insignificant essential data that can increasingly prompt accurate forecasts.

III. Proposed Methodology

In the recent year, deep learning methods become well known for their high accuracy rate and immense domain applications in various research fields, including image processing [24], [25], speech recognition [26], [27], computer vision [28], [29], authentication system [30]. Convolution neural networks (CNNs) feed-forward artificial neural network (FANN) expectant by standard procedures proposed to classify unique patterns straight away from medical and non-medical image data. Motivated by the great success of CNN in medical research, we employed CNN to target breast cancer tumor segmentation and classification. In convolutional neural networks, convolutional layers are the key building blocks used. The fast application of a variable to an input that results in inactivation is a convolution. Convolution of deep learning infrastructure focused on the neural network that performs medical imaging data analysis at different pipeline levels, including data loading, data increase network architectures, loss functions, and evaluation metrics.

In this paper, we present a new deep learning-based CNN framework architecture to segment and classify breast tumors into two classes (B-Benign and M-Malignant) by the use of CNN fine-tuned models. The proposed system consists of various pipeline stages, including data loading; patch extraction, selection, image segmentation, data augmentation, deep feature extraction, deep feature selection, and classification. The proposed system's detailed flow is shown in Fig. 2., whereas each pipeline stage's detail is described in subsequent sections. Models of deep learning are developed using neural networks. A neural network takes in inputs, and then processed in hidden layers using weights that are changed during preparation. The model spits out a forecast then. For making better predictions, the consequences are adapted to identify trends. A deep learning model is designed to continuously analyze data similar to how a person might conclude with a logic structure. A layered system of algorithms called deep learning applications to use an artificial neural network. The architecture of an artificial neural network is inspired by the human brain's biological neural network, leading to a learning mechanism that is far more capable than that of traditional models of machine learning.

A. Data Loading

In this stage, medical image files are loaded from a medical file format data set. Medical images are stored in different file formats as compared to many other computer vision tasks. These file format stores metadata information like acquisition information (specifies scanner parameter, modality types, etc.), spatial information (including anatomical point of reference and voxel anisotropy), and patients' data. Deep learning is a type of machine learning in which a model learns directly from pictures, text, or sound to perform classification tasks. Typically, deep learning is applied using the design of neural networks. The term deep refers to the number of layers in the network; the more profound the network, the more layers. The processing of medical imaging refers to the handling of images using a computer. This processing requires many methods and practices, such as image collection, storage, presentation, and communication.

B. Medical Image Segmentation

1. Patch Extraction

From the image file, we perform patch extraction over each image Im of size $P \times Q$; this image is then divided into patches Pch_i with a size of 256 × 256 pixels with no overlapping. For each image, the number of patches was different as the size of each image is different.

2. Patch Selection

Each image patches Pch_i are partitioned into sub-region SR_0, SR_1, SR_2, SR_3, SR_4 such that $\bigcup_{i=0}^{3} SR_i = Pch_i$ where ∪ representing the union

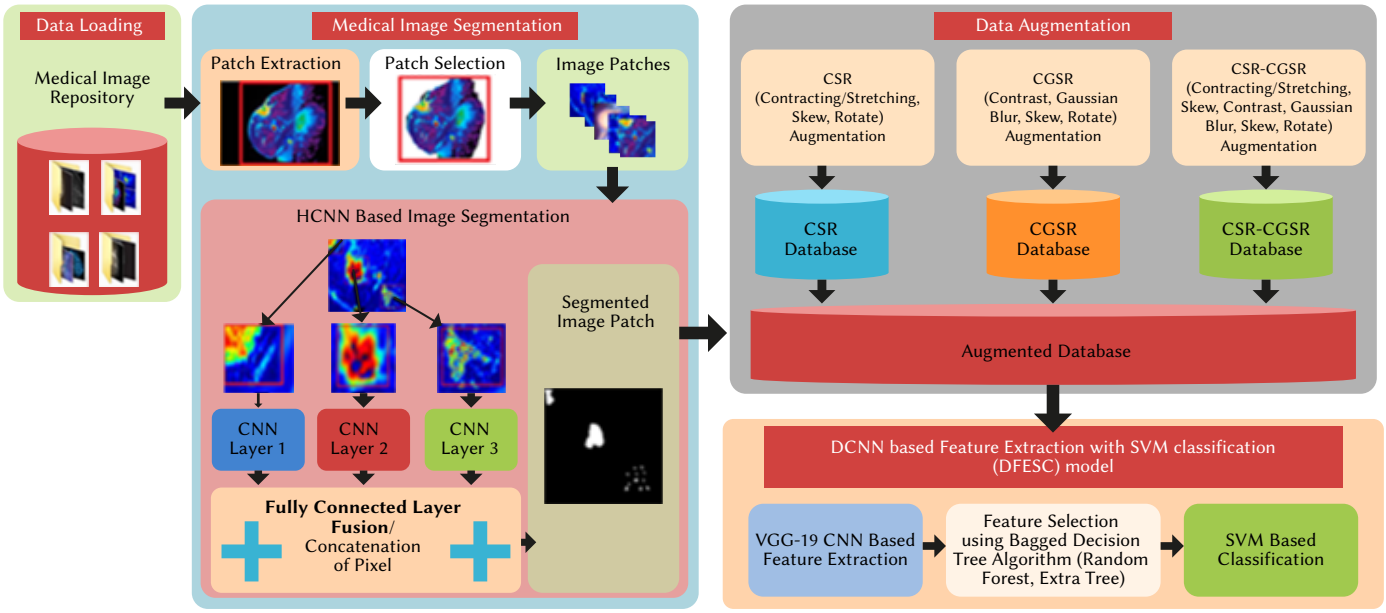

Fig. 2. Proposed Methodology for Tumor Segmentation.

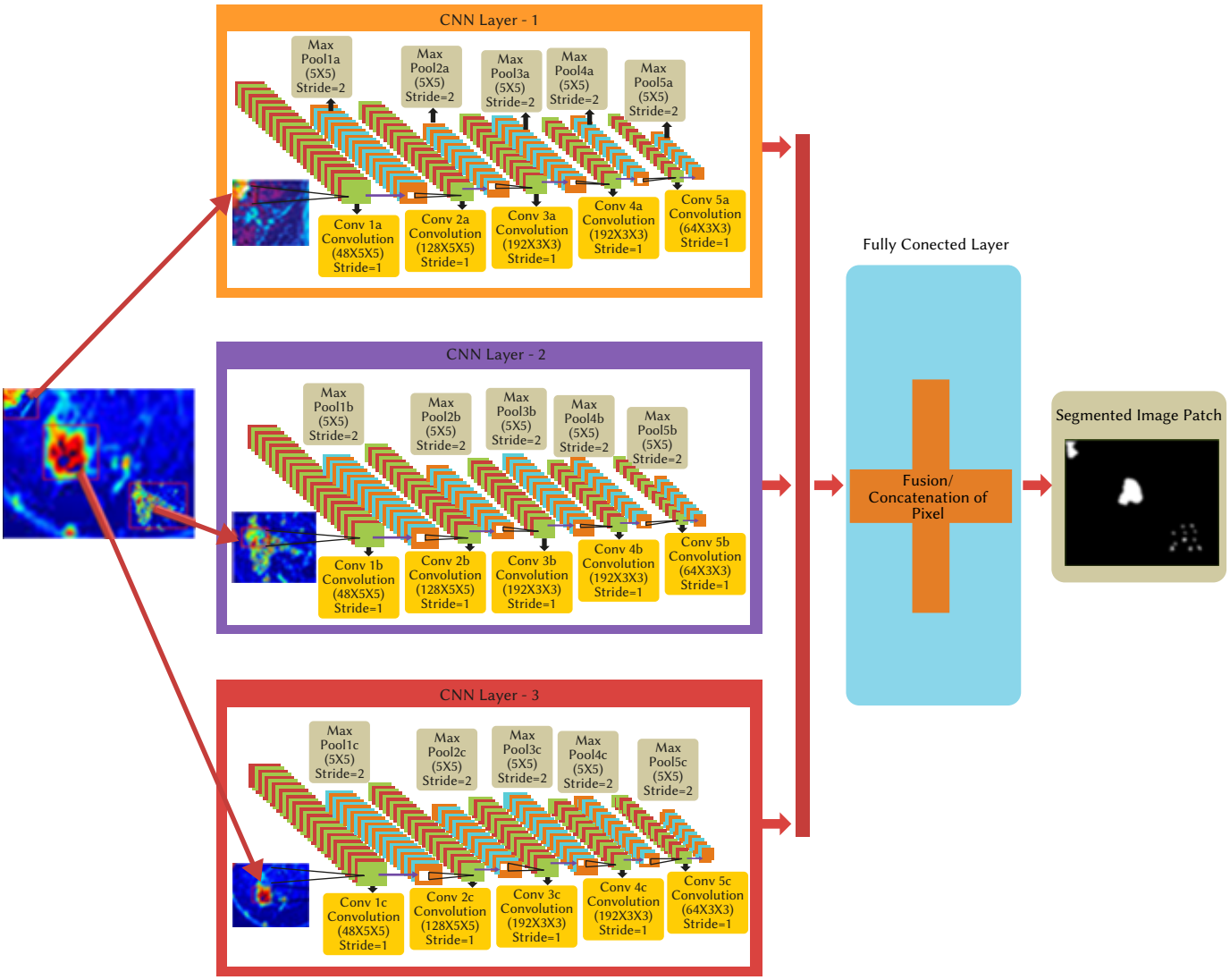

Fig. 3. HCNN architecture for Image Segmentation.

of all four regions. Only those regions are selected that contain tissue sections. Each patch has some noisy background data that is removed using the homogeneity factor, which is defined in (1):

$$H_f = \sum_{j=1}^{I} \quad \sum_{k=1}^{I} \quad \frac{Pr_{jk}}{1+|j-k|} \tag{1}$$

Where, Pr_{jk} is relative position probability of pixel pair (j, k), and I his distinct level of intensity was computed for every patch and optimized up to 60% threshold. Each patch belongs to the tissue region. Patching technology is an effective means of marking the structures of the brain and of other animals. In general, these approaches mark each of the voxels of a key growth by contrasting the image patch centered on the voxel with patches from the Atlas library. A search window is typically used locally based on the target voxel. Various patch-based mark fusion processes are proposed for effective and stable segmentation and illustrated. For works using non-rigid registration, comparable findings have been published. In a variety of computer vision tasks, including texture synthesis, painting and super-resolution, patch-based techniques recently showed high efficiency. Non-local denoising has led to the promotion of the field and to the development of many patch-based segmentation techniques for medical imaging applications.

The algorithm for patch extraction and patch selection is presented below:

Algorithm 1: Patch Extraction and Patch Selection

1. Extract the Patches Pch_i of size 224 × 224 pixel from $Im_{P\times Q}$
2. Partitioned patches into subregion SR_0, SR_1, SR_2, SR_3, SR_4.
3. Select Patches based on tissue section region such that for i=0 to 4:

   Begin:

   $\bigcup_{i=0}^{3} SR_i = Pch_i$

   End
6. Background noise removal using homogeneity factor for pixel pair (j, k).
7. For $j = 1$ *tol*:

   Begin: For $k = 1$ *tol*:

   Begin:

$$H_f = \sum_{j=1}^{I} \quad \sum_{k=1}^{I} \quad \frac{Pr_{jk}}{1+|j-k|}$$

   End

3. HCNN Based Image Segmentation

In the process of segmentation, background tissues are removed from the tumor region in the image. For segmentation, two methods are employed.

- Region-Based Approach in which segmentation is performed based on similarity detection. Few Region-Based approaches include Region growing, merging, and splitting using quad tree decomposition.
- Boundary Based Approach in which detection of discontinuity is performed and then linked to form boundaries of region.

A blend of distinct methods is implemented to maximize the segmentation outcomes. A region-based segmentation and picture analysis with application to medical images has been carried out in this paper. Clustering, object detection, and boundary detection are among the most critical measures in image segmentation. Segmentation of related structures is of utmost significance for several image processing and visualization activities both within and outside the medical image domain. As a result of non-optimal parameter settings, images segmented by area rising techniques often contain either too many regions or too few regions. A blend of distinct methods is implemented to maximize the segmentation outcomes.

The algorithm for boundary detection helps to find the right boundary for noisy pictures. The convergence between the original image and the corresponding mask provides insight into vector data. Finally, the algorithm for boundary detection is implemented to yield accurate input image boundaries.

In this paper, we perform Region of interest extraction using a deep learning-based fully automatic technique called Hierarchical CNN (HCNN). HCNN is different from traditional CNN because of its in-depth image processing. The architecture of HCNN consists of three hierarchical layers, which are fused at fully connected layers. Every pixel of the image is segmented, and the result of segmentation is then merged into the mask after input pixel segmentation. Fig. 3. represents the architecture of HCNN, including convolution layer, pooling layer, Rectified Linear Units Layer (RerLU), and fully connected layer.

Convolution Layer: It performs input image convolution using convolution kernels, which is represented in (2)

$$(R^k)_{x,y} = (Wt^k * Pch_i)_{x,y} + bs^k \tag{2}$$

Where * represent the convolution to the input image patches Pch_i, Wt^k and bs^k represent weight and bias between two neurons, whereas k represent convolution kernel index. $(R_k)_{xy}$ represent the convolution response between k^{th} kernel and pixel with center (x, y). To control the size of output volume of the convolution layer, we use three parameters: depth (number of convolution kernel), stride (control kernel shift amount), and padding (control the spatial size of convolution output volume).

Rectified Linear Units Layer (ReLU): It is used as an activation function that sets all the negative value to zero using the non-linear activation function shown in equation (3)

$$g(R) = max(0, R) \tag{3}$$

Where R represent convolution response output.

Pooling Layer: This layer performs non-linear transformation to reduce spatial dimension and noise elimination activated from the preceding layer. There are different down-sampling strategies used that perform pooling operations, including stochastic pooling, average pooling, and max pooling. Amongst all, Max pooling is most famous for its high-speed performance and convergence optimization.

Fully Connected Layer: It connects every neuron of the preceding layer to all neurons of this layer, called a fully connected layer. The (4) describes it:

$$(r^k)_{x,y} = \sum_f \quad (Wt^{kf} * Pch_i{}^f)_{x,y} + bs^k \tag{4}$$

Where, f represent f^{th} neuron index of input, r^k represent k^{th} neuron output, Wt^{kf} and bs^k represent weight and bias between two neurons Pch_i^f and r^k. In our proposed work, we concatenate the output of all the three convolutions at this layer. Health analytics analyses existing and past awareness of the market for predicting trends, increased scope and much better control the dissemination of diseases. It will include opportunities to enhance health safety, clinical data, diagnosis and organizational management. To overcome many technological challenges and issues that need to be solved to realize this opportunity, the motivational aspect of data analytics and mobile computing is critical for healthcare systems. New capabilities such as artificial learning, data analytics, and computational power have to be upgraded to provide more intelligent and skilled healthcare services for people in advanced healthcare systems.

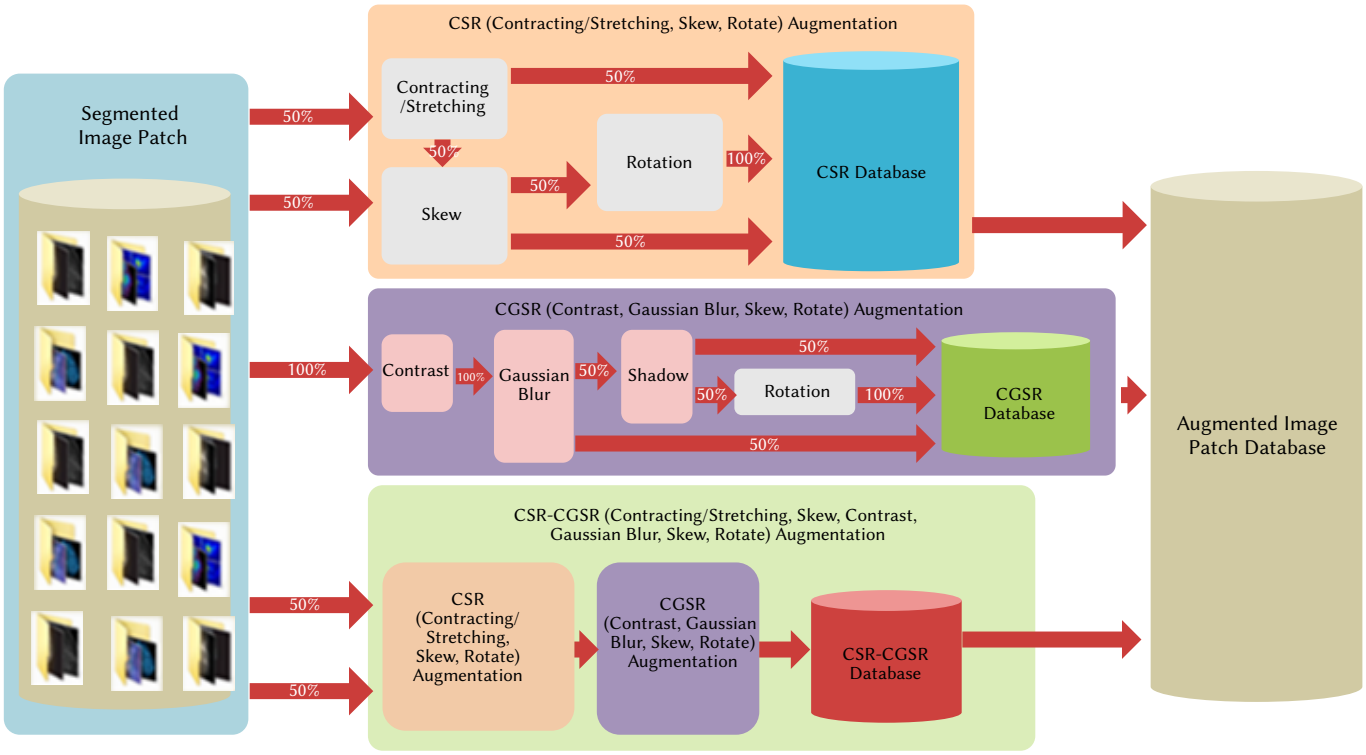

Fig. 4. Data Augmentation using CSR, CGSR, and CSR-CGSR Algorithms.

After performing these steps, a post-processing step is performed to eliminate the noise from the segmented image. The algorithm of the HCNN model for tumor segmentation is mention below:

Algorithm 2: Image Segmentation using HCNN model

1. Input: Image Patches Pch_i, Weight Wt^k and bias bs^k
2. Array Initialization: Number of filter ⟶ $NOF[5] = \{48, 128, 192, 192, 64\}$;
   Number of pixel in Imagepatches $Pch_i = 224 \times 224$;
2. Output: Segmented Image patches Pch_i'.
3. # Three Layer HCNN : for $i = 0$ to 2:
   # HCNNLayer 1: For $j = 0$ to 4
   Begin:
   Number of kernel: $NOK = NOF[j]$;
   if $(j < 2)$ then
   $Conv_{ij}$: $Conv(NOK, 5, 5, ImagePatch_{Size} = (Pch_i, Pch_i, 5))$
   $(R^k)_{x,y} = (Wt^k * Pch_i)_{x,y} + bs^k$
   else
   $Conv_{ij}$: $Conv(NOK, 3, 3, ImagePatch_{Size} = (Pch_i, Pch_i, 3))$
   $(R^k)_{x,y} = (Wt^k * Pch_i)_{x,y} + bs^k$
   $ReLu_{ij}$: $g(R) = max(0, R)$
   $MaxPool_{ij}$:
   $pool_{size} = (3, 3)$;
   $stride = (2, 2)$;
   End:
   End:
4. # *FullyconnectedLayer* 1: *Fusion of out put of all three HCNN layer*

$$(r^k)_{x,y} = \sum_f (Wt^{kf} * Pch_i{}^f)_{x,y} + bs^k$$

5. Output: segmented Image Pathes Pch_i'
6. Exit

C. Data Augmentation

This technique is used to increase the size of the training dataset and reduce overfitting. Data augmentation methods employed for geometric transformation invariance are rotation, shear, skewness, contracting/stretching, and flipping. In contrast, for noise invariance, the techniques used are edge detection, Gaussian blur, sharpen, shadow, and embossing. To make the training model robust and increase the training dataset, we propose three different combinations of augmentation: CSR (Contracting/Stretching, Skew, Rotate), CGSR (Contrast, Gaussian Blur, Skew, Rotate), CSR-CGSR (Contracting/ Stretching, Skew, Contrast, Gaussian Blur, Skew, Rotate). The complex transformation of Image patches is shown in Fig. 4.

D. DCNN Based Feature Extraction With SVM Classification (DFESC) Model

1. Feature Extraction

Feature extraction was carried out by passing the augmented image patches through the pre-trained fine-tuned VGG-19 network. DCNN based VGG-19 network is most popularly used because of its simplicity as it uses only three × three convolutional layers piled up on top of each other to increase depths. For dimensionality reduction, down-sampling of the input image (including image convolution, hidden-layer output matrix, etc.) is employed in this network. It consists of 19 layers, including five stages of 16 convolutional layers, rectified linear units (ReLU) activation, pooling layer of max type, and three fully connected (FC) layers. For better classification accuracy, the last FC layer is connected to the SVM classifier rather than the softmax layer that performs classification.

For feature extraction, first augmented image patches are normalized to zero mean and unit variance before feed into the VGG-19 network. The architecture VGG-19 based DFESC model is shown in Fig. 5 in which the first two layers of convolutional are trailed by max-pooling layer, and the same arrangement is continued for succeeding two layers as shown in Fig. 6. The remaining eight layers are arranged in a group of four convolutional layers followed by max pooling. This

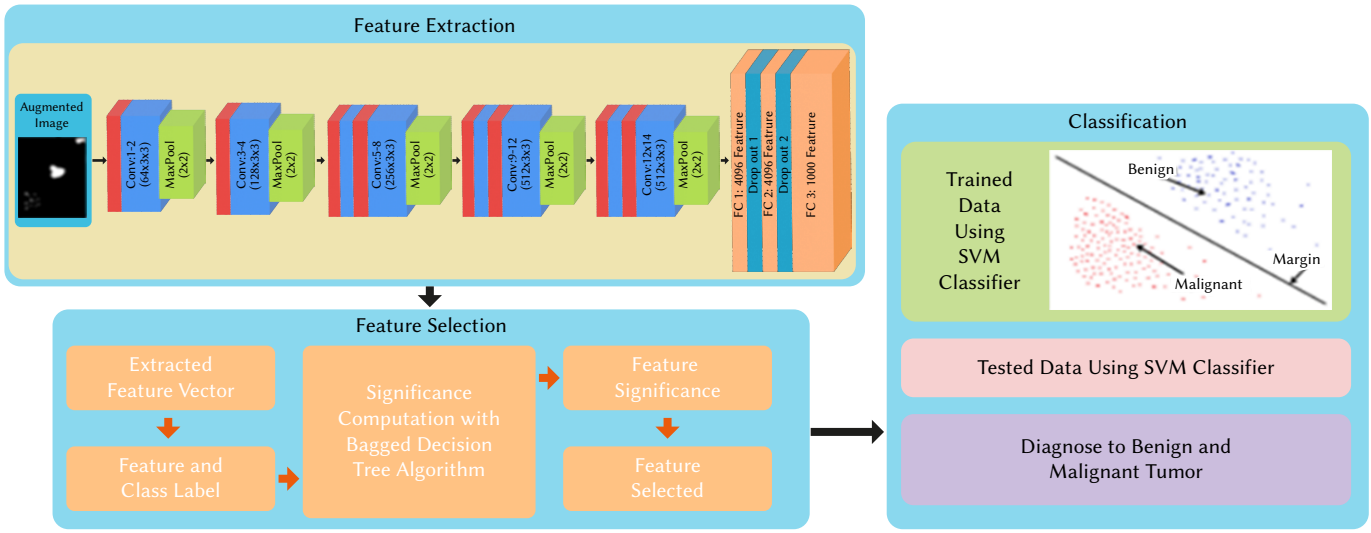

Fig. 5. Architecture of DCNN based Feature Extraction with SVM Classification (DFESC) Model.

Algorithm 3: Data Augmentation using CSR, CGSR, and CSR-CGSR Algorithm

- Input: Segmented Image Patches,
- Transformation parameter: Rotation (-90°, -45°,90°,45°), Gaussian Blur (σ=0.5,1.0,2.0.5), Contrast (lightness value =1.00, 1.5, 2.0, 0.25), Skew (Left, Right, Forward, Backward).
- Output: Augmented Image Patches
- Perform Data Augmentation using CSR, CGSR, and CSR-CGSR Transformation Algorithm
- CSR Augmentation Algorithm:
  i. First, split the segmented image patch database into two equal sets.
  ii. Apply Contracting/Stretching transformation to 50% segmented image patches, and the remaining 50% image patches are undergone through skew transformation.
  iii. Contracting/Stretching (CS) transformed images again split equally into two sets.
  iv. The applied skew transformation over 50% CS transformed image, and the rest of 50% CS transformed image is stored in CSR dataset.
  v. Skew transformed images again, split equally into two sets.
  vi. Applied rotation transformation over 50% skew transformed images, and the rest of 50% skew changed ideas is stored in the CSR dataset.
- CGSR Augmentation Algorithm:
  i. First, all segmented images are modified by the application of contrast transformation.
  ii. Transformed Contrasted images are then passed through Gaussian filters.
  iii. Shadow transformation is then further applied to images passed out from the Gaussian filter.
  iv. Shadow transformed images then split equally into two sets.
  v. Applied rotation transformation over 50% shadow transformed images, and the rest of the 50% shadow transformed image is stored in the CGSR dataset.
  vi. CSR-CGSR Augmentation Algorithm: Two augmentation techniques CSR and CGSR, are combined.
  vii. Initially, image transformation is performed according to the CSR transformation workflow, including Contracting/Stretching (CS), skewing, and rotation.
  viii. Afterward, CSR transformed images then go through CGSR transformation, including contrast reduction, Gaussian blurring, shadowing, and rotation.
  ix. All the transformed images are then stored in a CSR-CGSR dataset.

arrangement is then connected to the last three thoroughly combined (FC) that contains 4096, 4096, and 1000 nodes, respectively. The outcome from these layers resulted in 4096, 4096, and1000 features, respectively.

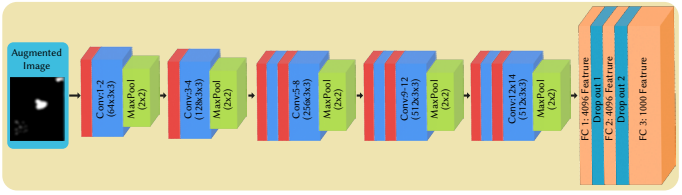

Fig. 6. Architecture of VGG-19 DCNN Based Feature Extraction.

2. Feature Selection

To reduce redundancy, we perform feature selection using bagged decision tree [31] algorithms like Extra Tree and Random Forest. Using these algorithms, we determine the significance of features and then select the quality based on their ranks using (5) in which a part with 95% significance is considered.

$$Significance\ of\ Selected\ Feature = 0.95 * Significance \quad (5)$$

3. Dataset Splitting

Selected feature vectors are then into three parts for training, validation, and testing. In this study, we hierarchically split the dataset. First, the dataset is divided into training and testing datasets in the percentage ratio of 85:15. The training dataset is then further divided into training and validation sets with a percentage ratio of 90:10. A diagrammatic view of the data split is shown in Fig. 7.

Fig. 7. Dataset Splitting.

4. SVM Classification

In this work, augmented feature vectors are combined with SVM classifiers that minimize classification error by determining the best possible separating hyperplanes. With a given labeled feature pair (p_f, q_f) where p_f represent feature vector and $q_f \in (+1, -1)$ represent

whether a given instance corresponds to the class or not. The (6) defines the formation of binary SVM cost function:

$$\frac{1}{2}\, x^T x + \mathrm{C}_r \sum_f \quad \varepsilon_f$$
$$subjecttoq_f \,.\, \left(x^T \emptyset\left(p_f\right) + \, y\right) \geq \, 1 - \varepsilon_f, \varepsilon_f \geq 0 \tag{6}$$

Where x and y represent separating hyperplane parameter, ε_f represent penalty error for a loose variable (located at the erroneous side of hyperplane margin), C_r represent control parameter for regularization that tradeoff between hyperplane margin and penalty error ε_f, $\emptyset(p_f)$ Represent the input vector non-linear transformation function. The two separatings hyperplane can be defined as $x^T \emptyset\ (p_f) + y = 1$ and $x^T \emptyset\ (p_f) + y = 0$, while margin width is defined as $\frac{2}{\|x\|}$. We use the radial basis function (RBF) kernel is used for SVM classifier, which is defined in (7):

$$K\left(p_f, p_g\right) \equiv \emptyset\left(p_f\right)^T \emptyset\left(p_g\right) = e^{-\gamma \|p_f - p_g\|^2} \tag{7}$$

Algorithm 4: DCNN based Feature Extraction with SVM classification

1. Input: AugmentedImage $Aug_I = 224 \times 224 \times 3$;
2. Array Initialization : Number of filter ⟶ $NOF_c[5] = \{64, 128, 256, 512\}$;
3. Feature extraction using VGG-19 net
4. f or $i = 1$ to 8 step 1

Number of kernel: $NOC_c = NOF_c\,[i]$;

Begin:

if $(i \leq 2)$ then

Begin:

for $j = 1$ to 2 step 1

Begin:

$conv_{(i,j)}$: $performconv(NOK_c, 3, 3;$ stride $= 1)$;

$(R^k)_{x,y} = (Wt^k * Aug_I)_{x,y} + bs^k$

$ReLu_{i,j}$:

$g(R) = max\ (0, R)$

End

$MaxPool_i$:

$pool_{size} = (2, 2)$;

$stride = (2, 2)$;

End

else if $(i \geq 3$ && $i \leq 5)$ then

Begin:

for $j = 1$ to 4 step 1

Begin:

$conv_{(i,j)}$: $performconv(NOK_c, 3, 3;$ stride $= 1)$;

$(R^k)_{x,y} = (Wt^k * Aug_I)_{x,y} + bs^k$

$ReLu_{i,j}$:

$g(R) = max\ (0, R)$

End

$MaxPool_i$:

$pool_{size} = (2, 2)$;

$stride = (2, 2)$;

End

else

Begin:

FC_i: Extract Feature

$ReLu_i$: Set all negative value to zero

$drop_i$: Perform 50% dropout

End:

End

5. Perform features election using bagged decision tree algorithm by estimating significance

Significance of Selected Feature = 0.95 * *Significance*

6. Perform SVM classifcation using RBF kernel

IV. Experimental Results

A. Dataset Description

DDSM [32] and CBIS – DDSM are the standard dataset containing a medical image of tumors for breast cancer detection and classification. CBIS – DDSM [33] is the latest version dataset that digitized mammogram images in DICOM standard format, while DDSM contains a lossless-JPEG design. These datasets were downloaded from the website of CBIS-DDSM that consist of 2478 mammography medical image of 1249 female, and these include mediolateral and craniocaudal (CC) oblique view. In this study, each oblique was considered as a separate image. The application of most of the sensors is now focused on evaluating the time domain of acquired sensors, typically by the magnitude and frequency of movement. However, this tentative analytical approach can neglect certain valuable sensing signals, such as identification information

We split the dataset images into three sets: a training set containing 1903 images, a validation set consisting of 199 photos, and a testing set with 376 illustrations. CBIS-DDSM dataset consists of pixel-based annotation for the Region of interest (ROI) with the label: Benign and Malignant. These labels are then further elaborated based on ROI as mass or calcification. Convert the downloaded mammogram images into PNG format and without cropping downsized images to 1152X896 using interpolation.

As we propose patch-based analysis, segmentation, and classification; therefore, we created two patch datasets by sampling patches of the image from the background and ROI region. We extract all image patches of 224×224-pixel size, and this patch size is sufficiently large enough that cover almost all ROI annotation. All patches were classified into one of the five categories: benign calcification, malignant calcification, soft mass, malignant mass, and background.

B. Performance Evaluation

In this section, we discuss the performance of the proposed system for multi-class breast cancer tumor classification. Therefore, performance evaluation of proposed DCNN based Feature Extraction with SVM classification (DFESC) model is discussed for two cases: Before Augmentation and After Segmentation. For the performance assessment of complex structures, there are many methods. However, most of the approaches proposed in the field of medical image processing only face the problem of identifying different metrics that allow precision from a strictly geometric and quantitative point of view to be assessed. Finally, potential future directions for performance assessment research in medical image segmentation are suggested. In essence, data integrity means that the data is correct and has not been wrongfully changed in any way. Inaccurate records can become a significant health concern for patients and an immense liability for providers, leading to fraud, misuse, lack of data, and incorrect or inadequate treatment.

In case 1, classification is performed over the original dataset (without augmentation), while in case 2, classification is performed over an augmented dataset, which is generated by using the proposed data augmentation algorithm (CSR, CGSR, CSR-CGSR). Classification result of proposed DFESC model with and without augmentation is shown in Fig. 8(a) and 8(b), whereas. Their corresponding confusion matrix analysis is shown in Fig. 9(a) and 9(b).

Fig. 8(a) and 8(b) showed that the proposed DFESC model with data augmentation algorithm predicted all five classes correctly with higher probability than without augmentation. It was also observed that background class is easily expected while malignant calcification class is hardest to predict.

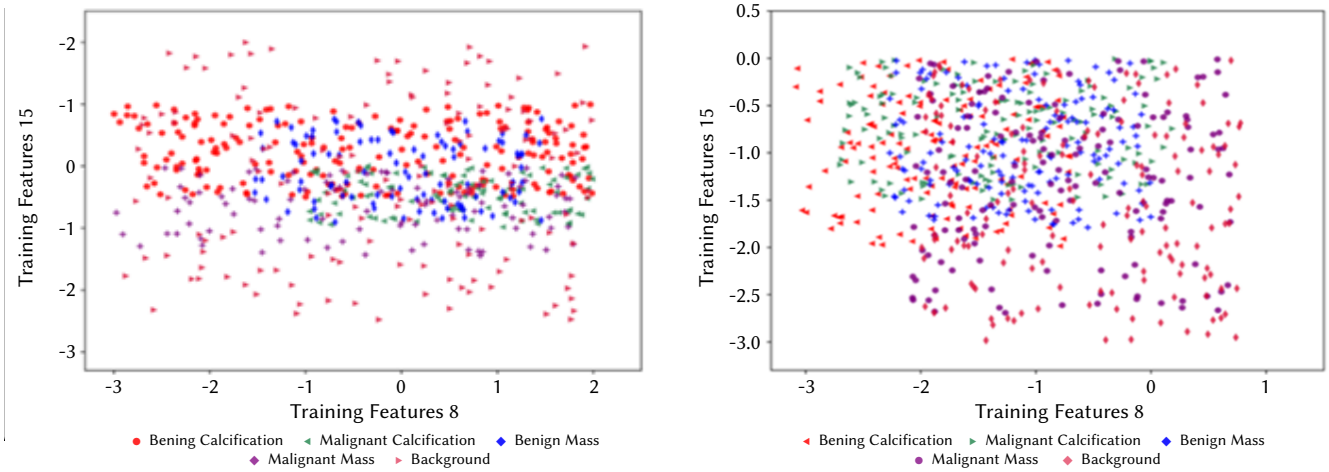

Fig. 8. Classification Result into 5-Class using Proposed DFESC Model (a) Without Augmentation (b) With Augmentation.

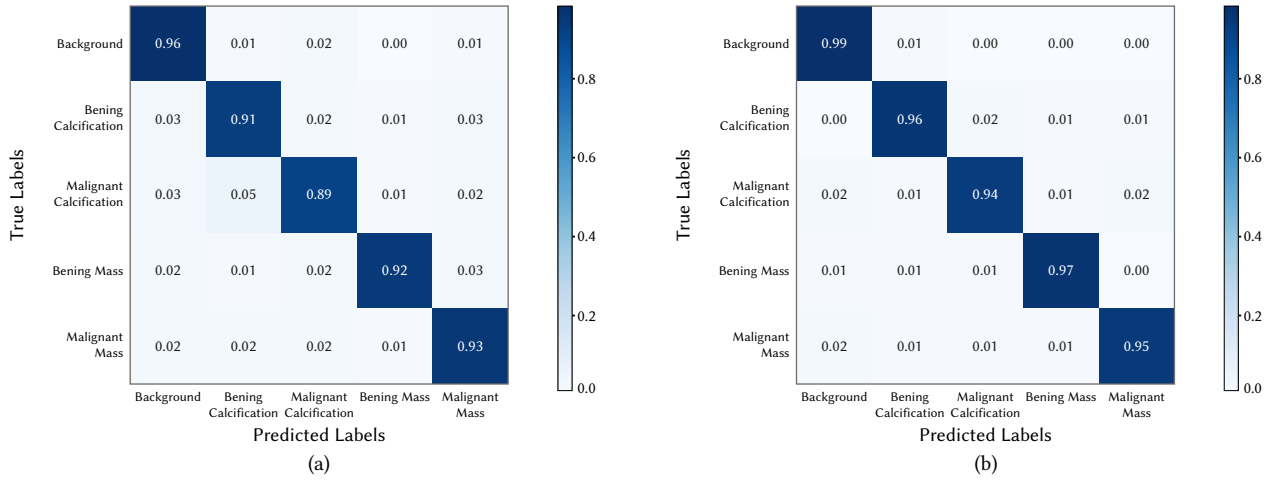

Fig. 9. Tumor Classification into 5-class using Confusion matrix analysis (a) Without Augmentation (b) With Augmentation.

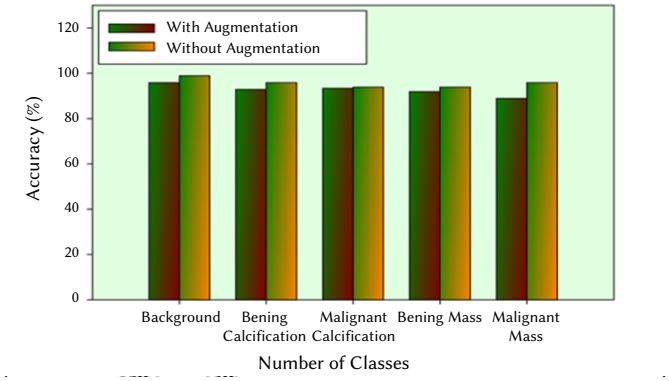

Fig. 10. Accuracy of each Tumor Class with and without Augmentation.

Based on the confusion matrix analysis, we formulate the proposed (DFESC) model's accuracy for both cases. Fig. 9 represents classification accuracy based on confusion matrix analysis for CBIS – DDSM dataset using the proposed classification model for both cases. Fig. 10 observed that classification accuracy after augmentation for background, benign calcification, malignant calcification, benign mass, and malignant mass are 99%, 96%, 94%, 97%, and 95%. In comparison, accuracy before augmentation is 96%, 91%, 93%, 92%, and 89% for background, benign calcification, malignant calcification, benign mass, and malignant mass. Based on this result, the data augmentation algorithm improves the classification accuracy up to a greater extent.

The overall classification accuracy with and without augmentation for the proposed DFESC model using the CBIS-DDSM dataset is shown in Fig. 11. It is marked from Fig. 11 that with the inclusion of the proposed data augmentation algorithm, the accuracy is improved largely. Accuracy starts from 40% for without augmentation while 58 % for the first epoch with augmentation dataset. In addressing these healthcare problems, big data analytics will help. With the help of predictive analytics, healthcare providers will cut healthcare costs and provide quality care. Big data frequently helps to minimize prescription mistakes by improving financial and administrative productivity and reducing hospital admissions.

Further, we assess the performance of the proposed work by computing AUCs per-image over an independent test set for two cases: Without Augmentation and With Augmentation. In case 1, classification is carried out over the original dataset (without augmentation), while classification is carried out over the augmented dataset in case 2, generated using the proposed data augmentation algorithm (CSR, CGSR, CSR-CGSR). The overall classification accuracy for the proposed DFESC model using the CBIS-DDSM dataset with and without augmentation is shown in Fig. 11. The accuracy is significantly enhanced with the inclusion of the submitted data augmentation algorithm.

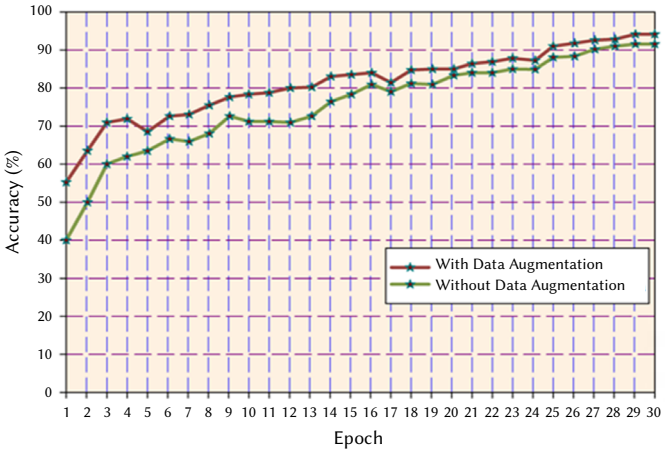

Fig. 11. Overall Accuracy of Proposed Method with and without Augmentation.

Case 1: Without Augmentation

First, we classify the tumor into 5- class over the original dataset (without augmentation) using the proposed DCNN based Feature Extraction with SVM classification (DFESC) model. Next, we generate a ROC curve with AUC computation for the DFESC model without performing augmentation. The ROC curve without augmentation into 5-class is shown in Fig. 12. The corresponding computed AUC values from the ROC curve are 0.91,0.98,0.97, 0.95, and 0.99 for benign calcification, malignant calcification, benign mass, malignant mass, and background. Overall average AUC of the proposed DFESC model before augmentation is 0.96.

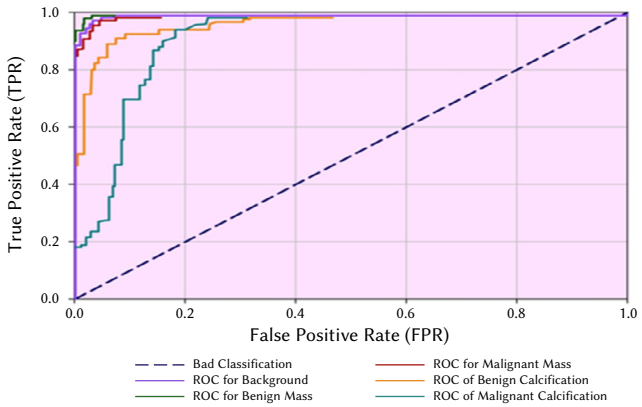

Fig. 12. ROC curve for Proposed DFESC model without augmentation.

Case 2: With Augmentation

In this case, we first augmented the original data using the proposed data augmentation algorithm (CSR, CGSR, CSR-CGSR) and then applied this augmented data to our proposed DCNN based Feature Extraction with SVM (DFESC) model for classification. Fig. 12 represents the ROC curve for benign calcification, malignant calcification, benign mass, malignant mass, and background class with an AUC of 0.990, 0.994, 0.991, 0.990, and 0.998 augmentations. Overall average AUC of the proposed DFESC model after augmentation is 0.994.

The ROC curve represented in Fig. 12 (without augmentation) and Fig. 12 (with augmentation) observed that classification accuracy for augmented data is higher for each class compared to without augmentation. Subsequently, compare the proposed DFESC model's performance using data augmentation algorithm with other CNN architecture, including ResNet, AlexNet, VGG-19, VGG-19 +ResNet. The performance comparison for all classification models using the ROC curve is shown in Fig. 13. It is evident from Fig. 14 that the proposed System represents a high AUC of 0.98 while the AUC of ResNet, AlexNet, VGG-19, VGG-19 +ResNet are 0.92, 0.88, 0.87, and 0.89. It shows that classification accuracy is high if the SVM classifier performs classification using RBF kernel over the feature extracted from the VGG-19 based DCNN.

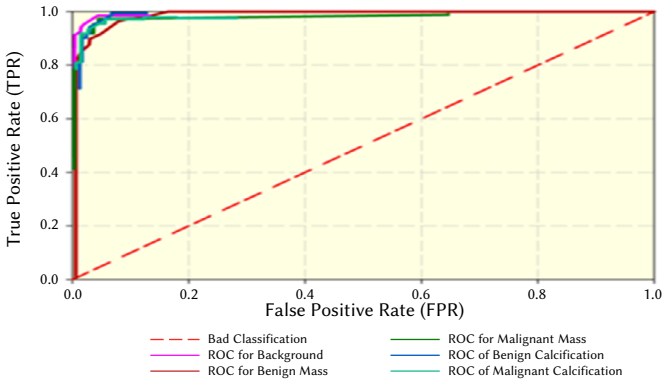

Fig. 13. ROC curve for Proposed DFESC model with Augmentation.

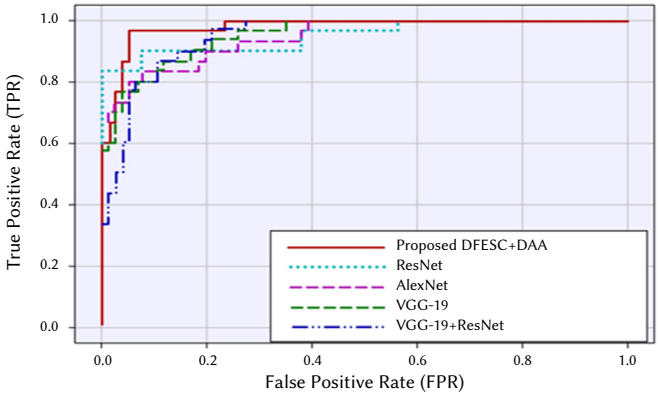

Fig. 14. ROC Curve Representation for Performance Comparison Analysis of Different CNN Classification Architecture with Proposed Method.

V. Conclusion

Deep Learning achieves a high level of accuracy in detecting and classifying multi-class breast cancer tumors. In this paper, we propose a combined approach based on deep learning and machine learning methods that perform classification over augmented data, resulting in better accuracy than other classification techniques. Authors suggest -fold method: Firstly, Authors employed Hierarchical CNN approach for tumor segmentation; Secondly, we propose three data augmentation algorithm-CSR, CGSR, and CSR-CGSR that perform augmentation over the segmented image; thirdly, we proposed VGG-19 CNN based feature extraction followed by feature selection using bagged decision algorithm (Random Forest, Extra Tree); finally, we perform classification using multi-class SVM classifier that classifies the breast cancer tumor image into five different classes: benign calcification, malignant calcification, benign mass, malignant mass, and background with high accuracy. Authors evaluate our proposed system's performance by performing the classification using the proposed DFESC model over original data and augmented data. From the simulation result, we observed that classifying the tumor using the DFESC model over augmented data achieves higher accuracy than without augmentation. Subsequently, the Authors perform a comparative analysis of different deep learning CNN architecture

(like ResNet, AlexNet, VGG-19, VGG-19+ResNet) with our proposed data augmentation based DFESC model. Form the comparison ROC curve, the Authors determine that the proposed method outperforms other CNN architecture. In the future, this work can be extended for fine-grained classification of each class with the examination of light-weight CNN architectures to steadiness the accuracy and efficiency.

References

[1] D. E. O'Leary, "Big Data, the Internet of Things and the Internet of Signs," *Intelligent Systems in Accounting, Finance and Management*, vol. 20, no. 1, pp. 53-65, 2013, doi: 10.1002/isaf.1336.

[2] D. W. Pennington, M. Margni, J. Payet, and O. Jolliet, "Risk and regulatory hazard-based toxicological effect indicators in life-cycle assessment (LCA)," *Human and Ecological Risk Assessment*, vol. 12, no. 3, pp. 450-475, 2006, doi: org/10.1080/10807030600561667.

[3] K. Van Hulle, "Solvency II: state of play and perspectives," *Zeitschriftfür die gesamte Versicherungswissenschaft*, vol. 100, no. 2, pp. 177-192, 2011, doi: org/10.1007/s12297-011-0138-2.

[4] H. K. Cho, K. P. Bowman, and G. R. North, "A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission," *Journal of Applied meteorology*, vol. 43, no. 11, pp. 1586-1597, 2004, doi: 10.1175/JAM2165.1.

[5] H. R. Boveiri, R. Khayami, M. Elhoseny, and M. Gunasekaran, "An efficient Swarm-Intelligence approach for task scheduling in cloud-based internet of things applications," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 9, pp. 3469-3479, 2019, doi: org/10.1007/s12652-018-1071-1.

[6] S. Thamburasa, S. Easwaramoorthy, K. Aravind, S. B. Bhushan, U. Moorthy, "Digital forensic analysis of cloud storage data in IDrive and Mega cloud drive," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2016, pp. 1-6.

[7] V. E. Sathishkumar, and Y. Cho, (2019, December). "Cardiovascular disease analysis and risk assessment using correlation based intelligent system," in *Basic & Clinical Pharmacology & Toxicology,* Hoboken, NJ USA, Wiley, 2019, pp. 61-61.

[8] A. Ahilan, G. Manogaran, C. Raja, S. Kadry, S. N. Kumar, C. A. Kumar, T. Jarin, K. Sujatha, M. K. Priyan, G. C. Babu, N. S. Murugan, and Parthasarathy, "Segmentation by fractional order darwinian particle swarm optimization based multilevel thresholding and improved lossless prediction based compression algorithm for medical images," *IEEE Access*, vol. 7, pp. 89570-89580, 2019, doi: 10.1109/ACCESS.2019.2891632.

[9] Y. Shi, C. Sun, Q. Li, L. Cui, H. Yu, and C. Miao, (2016, March). "A fraud resilient medical insurance claim system," in *Thirtieth AAAI Conference on Artificial Intelligence, (AAAI Press)*, Phoenix, Arizona, 2016, pp. 4393-4394.

[10] K. Fang, Y. Jiang, and M. Song, "Customer profitability forecasting using Big Data analytics: A case study of the insurance industry," *Computers & Industrial Engineering*, vol. 101, pp. 554-564, 2016, doi: org/10.1016/j.cie.2016.09.011.

[11] Y. Wang, and W. Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud," *Decision Support Systems*, vol. 105, 87-95, 2018, doi: org/10.1016/j.dss.2017.11.001.

[12] D. A. Koutsomitropoulos, A. K. Kalou, "A standards-based ontology and support for Big Data Analytics in the insurance industry," *ICT Express*, vol. 3, no. 2, pp. 57-61, 2017, doi: org/10.1016/j.icte.2017.05.007.

[13] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An ensemble random forest algorithm for insurance big data analysis," *IEEE Access*, vol. 5, pp. 16568-16575, 2017, doi: 10.1109/ACCESS.2017.2738069.

[14] Y. Ren, K. Zhang, and Y. Shi, (2019, November). "Survival Prediction from Longitudinal Health Insurance Data using Graph Pattern Mining," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, 2019, pp. 1104-1108.

[15] N. Rayan, "Framework for Analysis and Detection of Fraud in Health Insurance," in *2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Singapore, 2019, pp. 47-56.

[16] Y. M. Chae, S. H. Ho, K. W. Cho, D. H. Lee, and S. H. Ji, "Data mining approach to policy analysis in a health insurance domain," *International journal of medical informatics*, vol. 62, no. (2-3), pp. 103-111, 2001, doi: org/10.1016/S1386-5056(01)00154-X.

[17] M. S. Viveros, J. P. Nearhos, and M. J. Rothman, "Applying data mining techniques to a health insurance information system," in *VLDB'96, Proceedings of 22th International Conference on Very Large Data Bases*, Mumbai (Bombay), India, 1996, pp. 286-294.

[18] X. Jiang, S. Pan, G. Long, F. Xiong, J. Jiang and C. Zhang, "Cost-Sensitive Parallel Learning Framework for Insurance Intelligence Operation," in *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9713-9723, 2019, doi: 10.1109/TIE.2018.2873526.

[19] C. S. Wang, S. L. Lin, T. H. Chou, and B. Y. Li, B. Y. "An integrated data analytics process to optimize data governance of non-profit organization," *Computers in Human Behavior*, vol. 101, pp. 495-505, 2019, doi: 10.1016/j.chb.2018.10.015.

[20] N. S. Murugan, and G. U. Devi, "Feature extraction using LR-PCA hybridization on twitter data and classification accuracy using machine learning algorithms," *Cluster Computing*, vol. 22, no. 6, pp. 13965-13974, 2019, doi:10.1007/s10586-018-2158-3.

[21] N. S. Murugan, and G. U. Devi, "Detecting streaming of Twitter spam using hybrid method," *Wireless Personal Communications*, vol. 103, no. 2, pp. 1353-1374, 2018, doi: org/10.1007/s11277-018-5513-z.

[22] J. Sato, K. Goda, M. Kitsuregawa, N. Nakashima, and N. Mitsutake, "Novel Analytics Framework for Universal Healthcare Insurance Database," *AMIA Summits on Translational Science Proceedings*, vol. 2019, pp. 353-362, 2019.

[23] K. Umemoto, K. Goda, N. Mitsutake, and M. Kitsuregawa, "A Prescription Trend Analysis using Medical Insurance Claim Big Data," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, Macao, China, 2019, pp. 1928-1939.

[24] J. Ahmad, K. Muhammad, J. Lloret, and S. W. Baik, "Efficient conversion of deep features to compact binary codes using Fourier decomposition for multimedia big data," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3205-3215, 2018, doi: 10.1109/TII.2018.2800163.

[25] J. Ahmad, K. Muhammad, and S. W. Baik, "Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features," *Journal of medical systems*, vol. 42, no. 2, pp. 24, 2018, doi:10.1007/s10916-017-0875-4.

[26] J. Ahmad, K. Muhammad, S. I. Kwon, S. W. Baik, and S. Rho, "Dempster-Shafer fusion-based gender recognition for speech analysis applications," in *2016 International Conference on Platform Technology and Service (PlatCon)*, Jeju, Korea (South), 2016, pp. 1-4.

[27] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, "Efficient deep CNN-based fire detection and localization in video surveillance applications," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1419-1434, 2018, doi: 10.1109/TSMC.2018.2830099.

[28] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad and S. W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features," in *IEEE Access*, vol. 6, pp. 1155-1166, 2018, doi: 10.1109/ACCESS.2017.2778011.

[29] K. Muhammad, J. Ahmad, and S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *Neurocomputing*, vol. 288, pp. 30-42, 2018, doi: org/10.1016/j.neucom.2017.04.083.

[30] M. Sajjad, S. Khan, T. Hussain, K. Muhammad, A. K. Sangaiah, A. Castiglione, C. Esposito, and S. W. Baik, "CNN-based anti-spoofing two-tier multi-factor authentication system," *Pattern Recognition Letters*, vol. 126, pp. 123-131, 2019, doi: 10.1016/j.patrec.2018.02.015.

[31] D. Guan, W. Yuan, Y. K. Lee, K. Najeebullah, and M. K. Rasel, "A review of ensemble learning based feature selection," *IETE Technical Review*, vol. 31, no. 3, pp. 190-198, 2014, doi: org/10.1080/02564602.2014.906859.

[32] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammography," in *Digital mammography*, Springer, Dordrecht, 1998, pp. 457-460.

[33] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, pp. 170177, 2017.

Cheng Zhang

Cheng Zhang is an associate professor of the Department of Sociology in Jiangnan University, China. Her research interests include Social Security, Social Services, and Project Management, more than 9 papers published and 2 books published.

Dr. B. Vinodhini

Vinodhini. B. is presently an Assistant Professor in the Department of Computer Science & Engineering, SNS College of Technology, affiliated to Anna University-Coimbatore, Tamilnadu, India. Her research interests include Data Analytics, Computer Networks, and Wireless Communications. She filed 4 Patents, published 20 Research Articles and Published papers in International and National Conferences.

Dr. BalaAnand Muthu

Dr. BalaAnand Muthu is working as Associate Professor in the Department of Computer Science & Engineering at Adhiyamaan College of Engineering, India. His area of interest includes Big Data Analytics, Social Networks, Internet of Things in Healthcare. He is a member of IEEE and ACM. Has published many research articles in SCI, SCIE, Scopus indexed peer review journals. Also, handled Guest lectures, Intensive Workshop, Hands on programming in Hadoop, Spark, Grid & Cloud Computing at various technical institutions around Tamil Nadu. He is serving as reviewer in Computer Communication, IEEE Access, Multimedia Tools & Applications, International Journal of Parallel Programming, Enterprise Information System, Computer Networks, Measurement, Computer & Electrical Engineering, Wireless Personal Communication, Cluster Computing, Computational Intelligence, IET Transport Systems and so on.