Mohammed Anouar Naoui[1] / Brahim Lejdel[2] / Mouloud Ayad[3] / Riad Belkeiri[2] / Abd Sattar Khaouazm[2]

Integrating deep learning, social networks, and big data for healthcare system

[1] LIMPAF Laboratory, Computer Science Department, Faculty of Sciences and Applied Sciences, University of Bouira, Bouira, Algeria; El-Oued University, El-Oued, Algeria, E-mail: manouarn@yahoo.com
[2] Computer Science Department, El-Oued University, El-Oued, Algeria
[3] University of Bouira, Bouira, Algeria

Abstract:
This paper aims to propose a deep learning model based on big data for the healthcare system to predict social network data. Social network users post large amounts of healthcare information on a daily basis and at the same time hospitals and medical laboratories store very large amounts of healthcare data, such as X-rays. The authors provide an architecture that can integrate deep learning, social networks, and big data. Deep learning is one of the most challenging areas of research and is becoming increasingly popular in the health sector. It uses deep analysis to extract knowledge with optimum precision. The proposed architecture consists of three layers: the deep learning layer, the big data layer, and the social networks layer. The big data layer includes data for health care, such as X-ray images. For the deep learning layer, three Convolution Neuronal Network models are proposed for X-ray image classification. As a result, social network layer users can access the proposed system to predict their X-ray image posts.

Keywords: big data, deep learning, healthcare system, social network, X-ray image

DOI: 10.1515/bams-2019-0043

Received: August 16, 2019; **Accepted:** December 23, 2019

Introduction

Health systems have emerged as an area of interest for research, particularly with regard to the large amounts of heterogeneous healthcare data. In recent years, healthcare data has become more and more complex because there are many types of data, such as structured, semi-structured, and unstructured data.

There are also multiple sources of healthcare data, such as social networks, laboratories, medication, nurses' instruments, etc. As a result, the volume of information stored in the health system is constantly increasing and traditional software has difficulty to manage it.

Big data architecture can improve the healthcare system by providing an ideal framework for data, referring to complex and voluminous data that can easily manage voluminous data from multiple sources and offering a diversity of data types.

Several studies have focused on the use of social networks in the health system, which they have mobilized for various reasons [1]:

- Recruitment for clinical trials;
- Professional development and training for clinicians;
- Inter-professional communication and coordination;
- Health advocacy and fundraising for health organizations;
- Development of interactive, self-management tools and plugin to popular social media platforms;
- Public health messaging;
- Infectious disease monitoring.

Mohammed Anouar Naoui is the corresponding author.

This work proposes a deep learning model for healthcare prediction problems in social networks. The architecture of this model combines deep learning, big data, and social networks. In addition, the proposed method aims to answer the following questions:

How to store data important to the health system in a distributed and heterogeneous system?

- What is the model that can describe health system data?
- How can social network users predict its data, for example, an X-ray image?

The remainder of the paper is structured as follows: in Section Related work, related papers that outline several types of research in the social network, healthcare, and machine learning are presented. Section Proposition presents the proposed approach. Section Validation presents the validation of the system. The results are discussed in Section Results and discussion. Finally, Section Conclusion and future work presents a brief conclusion of this research study and some perspectives for future work.

Related work

Social networks and sites for the healthcare sector

Nowadays, there are several social networks and sites dedicated to the health sector. Figure 1 [2] is a social networking site where users can share their knowledge; for example, users can post an X-ray image and other users can comment on this image. Open-iBiomedical [3] Image is a site containing thousands of medical images with all the information about the disease and the patient's age, condition, diagnosis, and treatment. It is one of the most important sites used to manipulate text and images.

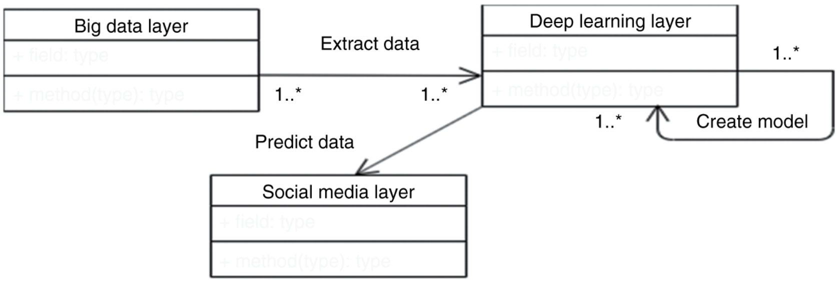

Figure 1: Unified modeling language class diagram of the proposed system.

MedPix [4] is a free online database with open access to medical images, clinical subjects, image merging, and textual metadata including over 12,000 patients case scenarios, 9000 themes, and approximately 59,000 images. Its primary target audience includes doctors, nurses, medical students, and nursing students.

Social network data for healthcare prediction

Several studies have examined data collected from social networks disease prediction, atherosclerotic heart disease [5], public health [6], or the infectious disease spread [7]. These researches collect data from social networks to build a model that can predict diseases.

Data mining for healthcare

Some research surveys have addressed the challenges of data mining in healthcare [8], [9], [10], [11], while others have focused on disease prediction [12], [13], [14]. In addition, several studies have proposed models and frameworks for healthcare system, such as the decision tree [15], [16], [17], [18], the Support Vector Machine [19], [20], [21], [22], Neuronal networks [23], [24], [25], the Bayesian method [16], [26], Regression [27], [28], Clustering [29], [30], [31], [32], [33], [34], or the Apriori algorithm [35], [36].

Moreover, several authors have highlighted the benefits of data mining in the health sector. These include the availability of medical solutions for patients at lower cost; the detection and prediction of the cause of diseases and the identification of medical treatment methods; and effective decision-making support for health researchers [8].

Big data for healthcare

The characteristics of big data are known as 5V namely volume, variety, velocity, veracity, and value. Thus, a model based on big data is an ideal model for the healthcare system. As a result, several studies have discussed the volume of data in the healthcare system such as personal information, radiology images, 3D imaging, genomics, signal processing, and biometric sensor reading [37], [38]. There are a variety of data types: structured, semi-structured, and unstructured. Velocity refers to the speed at which data are generated, for example, X-ray data. Veracity in healthcare system means that the big data system provides certificate information on diagnostics/treatment. Value refers to the quality of the information.

Deep learning in healthcare

Traditional data mining methods tend to favor simple and structured data, which may not be able to effectively use the rich information in healthcare data. The latest paradigm in deep learning technologies allows information to be extracted from complex and heterogeneous sources in an efficient way. Several researches in deep learning for healthcare have proposed a model for learning healthcare data cancer diagnostics [39], drug design [40], disease prediction [41], human behavior monitoring [41], and lifestyle-related diseases [42].

Proposition

Architecture of the proposed system

Figure 1 presents the architecture of the proposed system. It is composed of three layers: big data layer, deep learning layer, and social network layer.

- Big data layer: this layer is responsible for storing and managing health data. It provides users with a flexible framework for data that is characterized by volume, variety, velocity, veracity, and value.
- Deep learning layer: this layer offers users multiple deep learning algorithms. It analyzes the data stored in the big data layer to provide deep results to system users.
- Social network layer: it is responsible for social networks as a source of healthcare information. There are several social networks that can be classified into two categories according to healthcare information, general social networks: containing general information including healthcare, for example, Facebook, Twitter, etc.; and healthcare social networks: that only include healthcare information, for example, Figure 1 [2].

Figure 2 presents the components of the proposed architecture. It is composed of:

- HDFS File: Hadoop Distributed File System, which is a distributed file system used to store very large dataset running on clusters reliably. HDFS provides fast and automatic failure detection/recovery, supports continuous access to datasets, facilitates the processing of large datasets and data portability from the heterogeneous platform.
- Data i: 1.n: It represents data stored in big datasystem, data of healthcare from hospital clinical data, etc.
- Deep learning algorithms i: 1.n: Refer to the worker run on data, which are deep learning algorithm tasks, the model i:1.n: the trained model built with deep learning algorithm.
- Social networks: Social media users connect with the model to predict its data, for example, a classification of an X-ray image.

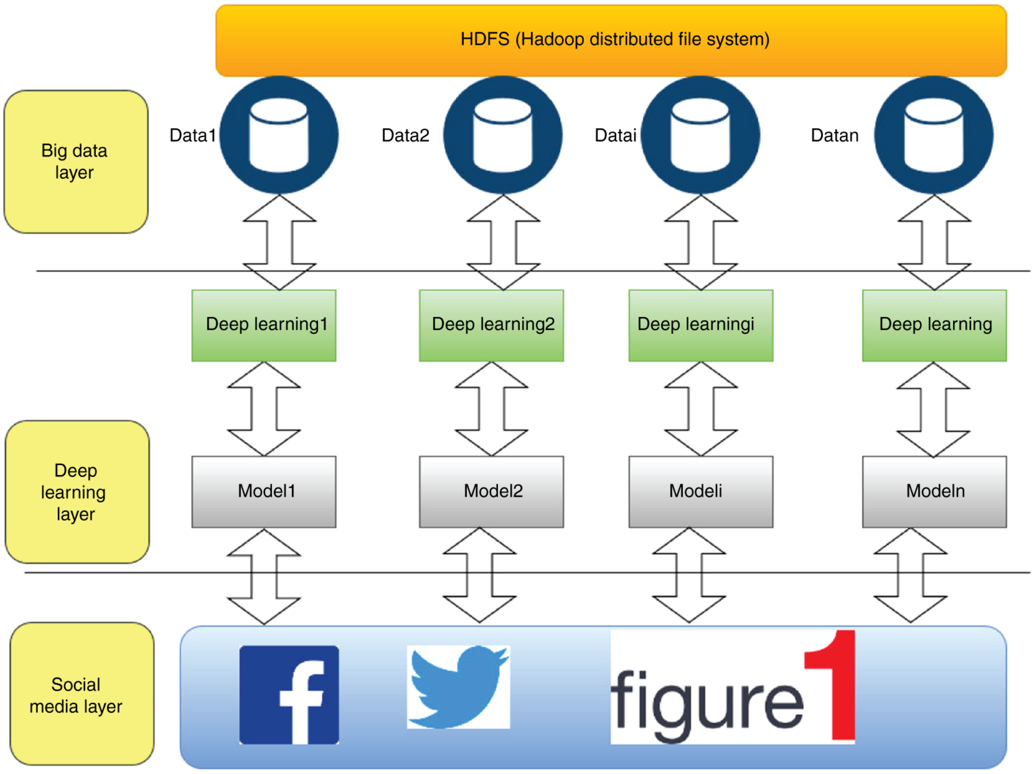

Figure 2: Big data, deep learning, and social network layers of the proposed system.

Use case (X-ray image prediction)

This section presents the case study of an X-ray image, its dataset and several deep learning algorithms based on Convolution Neuronal Network (CNN) to predict an X-ray image. The CNN algorithm is one of the most important algorithms in image classification, therefore it is applied to classify diseases using X-ray images. There are multiple and large image sources, therefore the big data tools are proposed to store and manage them.

Researchers have identified many problems in the diagnosis of X-ray images, either because of the diversity of diagnoses or the difficulty of the diagnosis itself. As a result, deep learning can help to address these challenges as traditional algorithms are no longer sufficient to solve them. It is also a valuable complement in the medical field and helps doctors and their patients as well.

In addition, deep learning can help users solve the confusion of the wrong diagnosis exists in X-ray images. The purpose of studying X-ray images is to analyze the huge dataset in healthcare system. For example, the NIH Institute published thousands of images in October 2017.

Moreover, the numerous chest diseases and the great convergence between them makes it difficult for a noncompetent person to distinguish them. Thus, this study uses X-ray images from Chest X-ray14, Shenzhen Hospital, and the Montgomery County X-ray dataset.

Dataset description

The National Institutes of Health (NIH, which can be translated by U.S. Institutes of Health) are government institutions in the United States that are involved in medical and biomedical research. They depend on the Department of Health and Social Services of the United States [43].

On October 2017, the National Institute of Health opened the following sources of chest X-ray images named Chest X-ray14. The photos were released to allow doctors to make better diagnostic decisions for patients with lung disease 112.120 chest X-ray PNG images in 1024 × 1024 resolutions [43]. Chest X-ray14 includes 14 diseases: atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, pleural thickening, consolidation, emphysema, hernia, fibrosis, and edema.

Shenzhen Hospital dataset: X-ray images in this data collection have been collected by Shenzhen Hospital, in Guangdong Province from PR China. X-ray images have been obtained as part of routine care at Shenzhen

Hospital. The collection contains JPEG images. There are 340 normal X-ray images and 275 abnormal X-ray images [44]. Shenzhen data is divided into two classes: normal and abnormal.

Montgomery County X-ray Set: X-ray images in this data collection have been collected by Shenzhen Hospital, in Guangdong Province from PR China. X-ray images have been obtained as part of routine care at Shenzhen Hospital. The collection contains JPEG images. There are 340 normal X-ray images and 275 abnormal X-ray images [44]. Montgomery data is also divided into two classes: normal and abnormal.

Big data layer

In the big data layer, images are stored in the HDFS nodes. The deep learning trains on data to create a model. Figure 3 shows the components of big data layer.

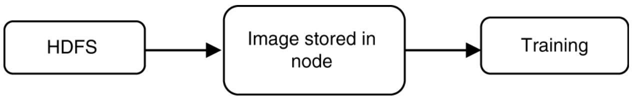

Figure 3: Image data based on Hadoop Distributed File System.

Deep learning layer

Recently, CNN models have been very successful in recognizing and classifying images in large datasets due to their ability to use highly scalable training algorithms. These algorithms are used to provide deep learning for images in social networks. Therefore, deep learning can be used to analyze MRI and X-ray images, to help diagnose and make a prognosis.

A CNN is a feed-forward network and is composed of three layers: convolution layer, pooling layer, and fully connection layer. The convolution layer extracts the feature representations from their input images. The neurons in this layer are adapted into feature maps. Each neuron has a receptive field, connected to a neighborhood of neurons in the previous layer via a set of trainable weights.

The role of the pooling layer is to reduce the spatial resolution of the feature maps, therefore achieve spatial invariance to input distortions and translations [45]. Each neuron in the previous layer is connected with the next neuron in the next layer. Finally, the objective of the fully connected layer is to classify the input image based on the training dataset.

Table 1 summarizes three proposed CNN models for classification problems.

Table 1: Convolution Neuronal Network (CNN) models.

Models	Classification
CNN1	Chest X-ray images/not chest X-ray images/other X-ray images
CNN2	Normal/Abnormal
CNN3	Disease classification

CNN1

This model is the simplest classification model. The difference between the images is easy to identify. The input is an image and the output is a class of the image, whether it is a chest X-ray image, not chest X-ray image,or other X-ray image. Six layers of neural network are proposed: 2 convolution layers, 2 max pooling layers, and 2 fully connected layers with 32 filter matrices. The ReLu activation function is used at the end of neural with learning rate of 10e-4. Figure 4 presents CNN1 architecture model.

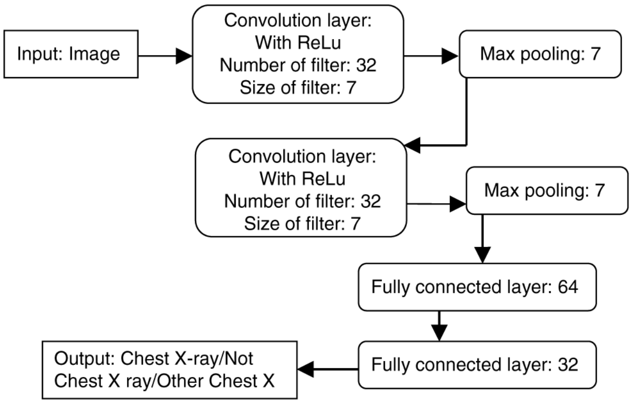

Figure 4: Convolution Neuronal Network (CNN1) architecture model.

CNN2

For this model, the input is an X-ray image and the output is the health status of the image: normal or abnormal. This CNN model consists of a larger number of layers than the previous algorithm: 3 convolution layers, 3 average pooling layers, and 3 fully connected layers with an increase in the number of filters (up to 128 filters).

The model was tested a first time at a 10e-4 learning rate and a second time at 10e-5 learning rate. Figure 5 presents the CNN2 architecture model.

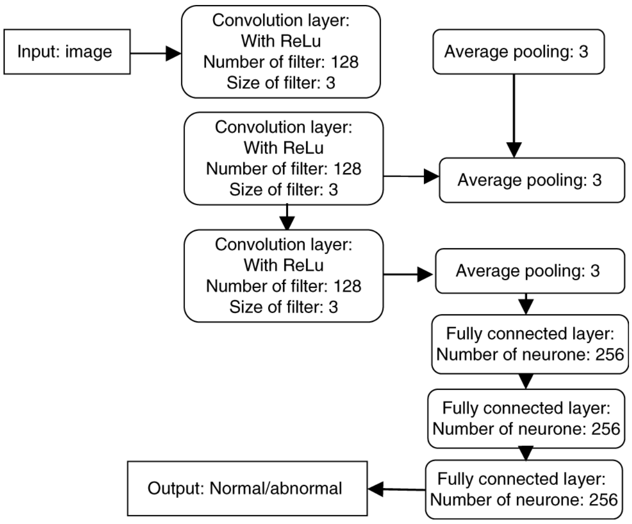

Figure 5: CNN2 architecture model.

CNN3

This is the most difficult model and it requires a larger number of layers and filters and reduces the learning rate with increasing the number of epochs. The input is an X-ray image of a patient's condition (abnormal) and the output is the name of the disease. It has 9 layers and 128 filters, for a learning rate of 10e-5, using the ReLu function. Figure 6 presents the CNN3 architecture model.

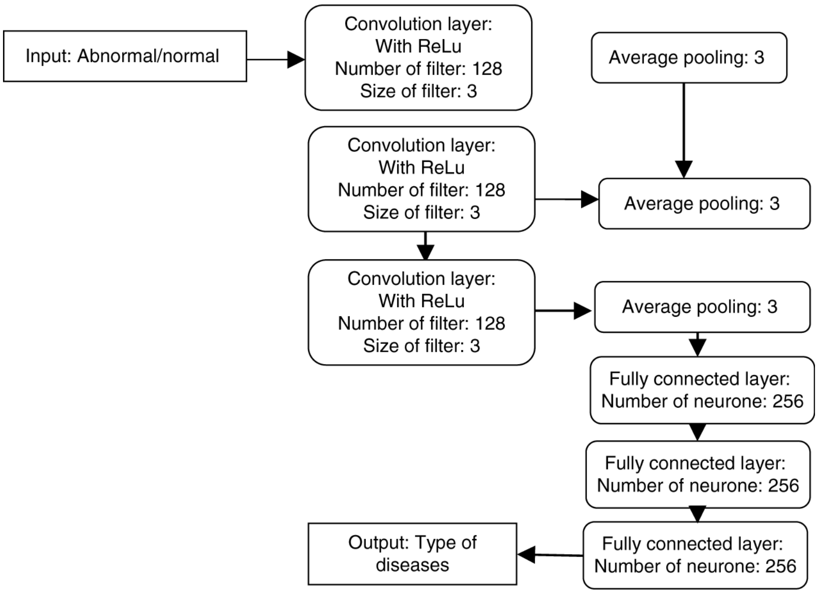

Figure 6: CNN3 architecture model.

Social media layer

Twitter is one of the most popular social network sites. For this study, the system downloaded images published in the tweets posts. Then, the model was tested to replay Twitter comment for image classification. Figure 7 shows social media layer predict process.

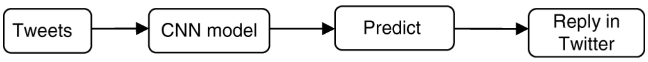

Figure 7: Social media layer predict process.

Validation

The proposed model was validated with Python Hadoop and TensorFlow. Python has access to high-quality libraries; therefore, it is the most used in the field of artificial intelligence and machine learning. TensorFlow, Tweepy, and Pyspark are all supported by Python. Its interpreters are available for more operating systems.

The most well-known libraries in the field of artificial intelligence are built-in Python, such as TensorFlow, Scikit-learn, and Keras. TensorFlow was created to use Python language, although there are other programming languages used in these libraries, but not as effective as Python [46].

Hadoop is an open-source software platform written in Java to store and manipulate massive data in a distributed format such as storing huge data on multiple devices and then distributing the processing process to these devices to speed up the processing result [47].

TensorFlow, developed by Google and released in 2015, is an open-source machine learning framework for high-performance numerical computation. Many companies currently use TensorFlow such as IBM, Twitter, Intel, Nvidia [48].

In this study, TensorFlow was used for three CNN algorithms to determine the type of image (chest X-ray, not chest X-ray, other X-ray); to determine if someone is infected or not (normal or abnormal); and to identify the type of disease.

First, CNN1 was trained to determine the type of image (chest X-ray, not chest X-ray, other X-ray). Then, CNN2 was trained with chest X-ray photos of the datasets to determine if the person is infected or not (normal or abnormal). Finally, CNN3 was trained based on the NIH Institute dataset that contains approximately 112,000 X-ray images divided into 14 classes. Thus, TensorFlow was used to identify the type of disease. Figure 8 presents the image prediction process.

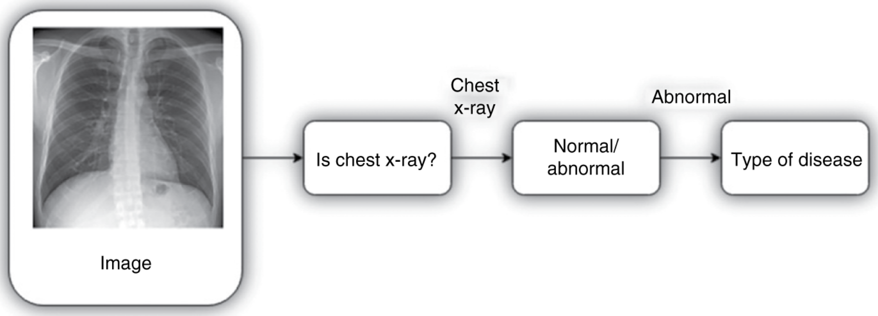

Figure 8: Image prediction process with the proposed system.

In addition, the proposed system shows how to obtain images from Twitter via Twitter API developed by the same company that provides full access to tweets and images. Once someone publishes a chest X-ray image, it is recorded, its infection rate is obtained, the type of disease is determined, and the tweet is then answered with a comment. The images are stored in several nodes designed by Hadoop, each one contains a wide range of images, and the proposed system will access to these images and train them to obtain the model.

Table 2 summarizes the parameters used for each CNN model. Figure 9 draws the tensor board of neuronal network for CNN1 and Figure 10 for CNN2.

Table 2: Parameters of CNN 1, 2, and 3.

CNN	CNN1	CNN2	CNN3
Batch size	16	16	16
Image size	256	300	300
Number of files in training-set	2500	2500	112,000
Number of layers	6	9	9
Number of files in validation-set	600	500	12,000
Number of epoch	30	20	20
Learning rate	10e-4	10e-4	10e-5
Activation function	ReLu	ReLu	ReLu
Accuracy validation	95%	65%	70%

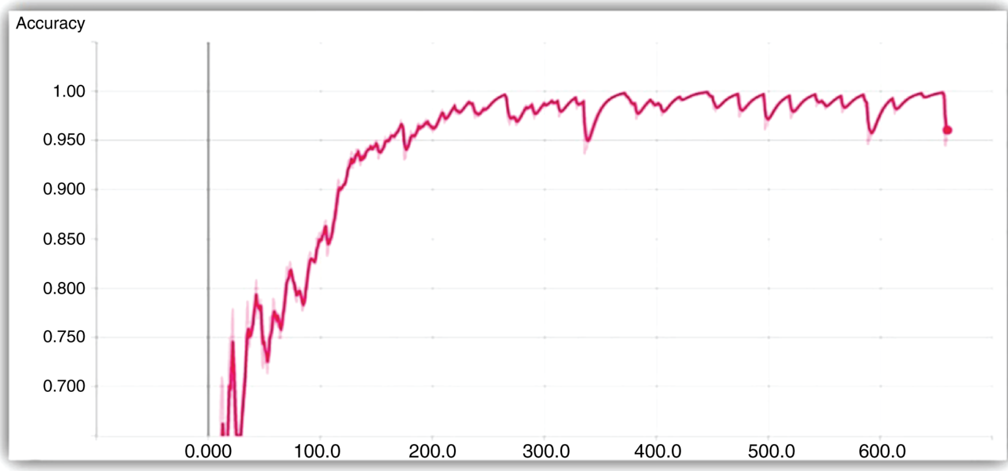

Figure 9: TensorBoard training accuracy of CNN1classification.

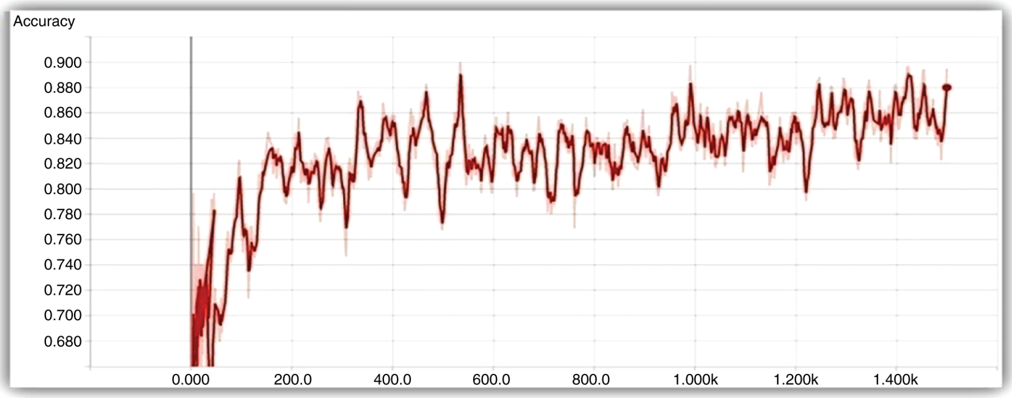

Figure 10: TensorBoard training accuracy of CNN2 classification.

Figure 11 shows three different images that were used to test the result of CNN1. In the first image, the accuracy value of the chest X-ray image is 99.519%, for the second image (not chest X-ray) is 99.97%, and for the third image (other X-ray) is 99.03%.

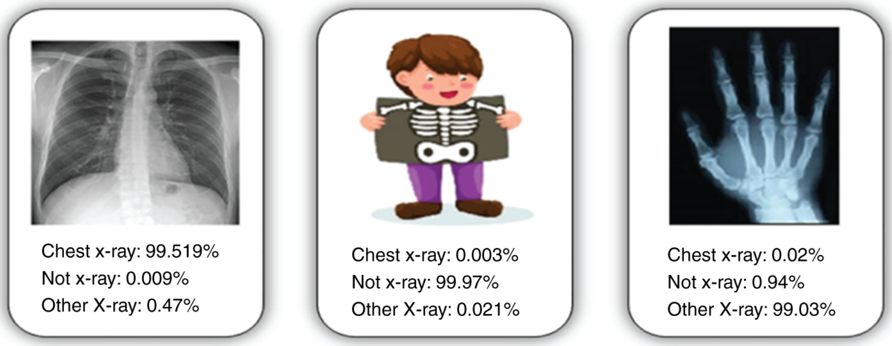

Figure 11: Image testing for CNN1 classification based on three different images.

Figure 12 shows the classification of CNN2 for the test of image return 67% for normal class.

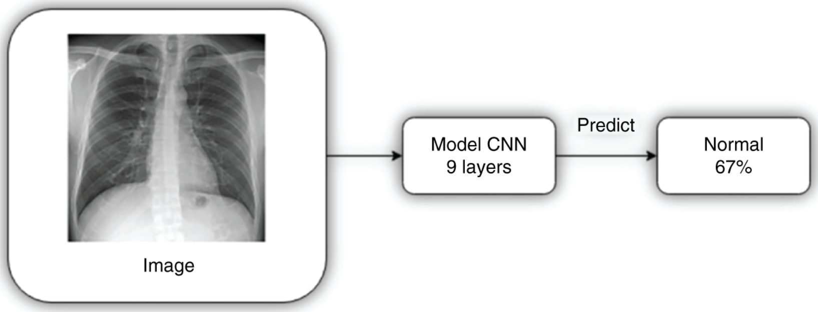

Figure 12: Image testing for CNN2 classification.

The input in Figure 13 is an abnormal X-ray image and the result is pneumonia with 62%.

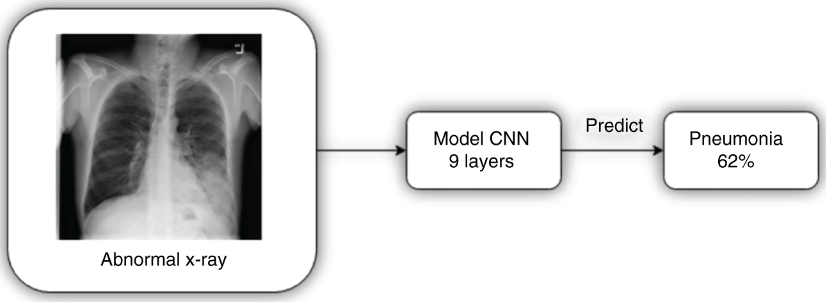

Figure 13: Use of CNN3 model for the prediction of pneumonia disease.

Table 3 provides the accuracy rate of the CNN3 model for each disease.

Table 3: Accuracy rate of each disease with CNN3 model.

Abnormality	Accuracy
Atelectasis	0.758
Cardiomegaly	0.786
Effusion	0.65
Infiltration	0.8
Mass	0.7
Nodule	0.756
Pneumonia	0.62
Pneumothorax	0.795
Pleural thickening	0.75
Consolidation	0.645
Empyshema	0.72
Hernia	0.62
Fibrosis	0.64
Edema	0.69

Twitter user's X-ray image prediction

A Twitter user posts an X-ray image and the proposed model can classify it in the user comments. Figure 14 shows the proposed model comment for abnormal (76%) and normal (24%).

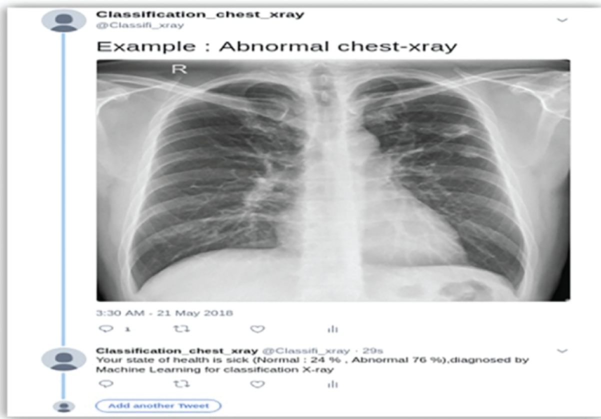

Figure 14: Comments based on the CNN2 model image classification.

Results and discussion

The proposed model was compared with the model proposed by Wang et al. [43], AlexNet, GoogLeNet, VGGNet-16, and ResNet-50. Table 4 summarizes the results obtained for eight abnormalities.

Table 4: Comparison between the proposed model and Wang et al. [43], AlexNet, GoogLeNet, VGGNet-16, and ResNet-50.

Abnormality	Wang et al.	AlexNet	GoogLeNet	VGGNet-16	ResNet-50	The proposed model
Atelectasis	0.707	0.6458	0.6307	0.6281	0.7069	0.758
cardiomegaly	0.8141	0.6925	0.7056	0.7084	0.8141	0.786
Effusion	0.7362	0.6642	0.6876	0.6502	0.7362	0.65
Infiltration	0.6128	0.6041	0.6088	0.5896	0.6128	0.8
Mass	0.5609	0.5644	0.5363	0.5103	0.5609	0.7
Nodule	0.7164	0.6487	0.5579	0.6556	0.7164	0.756

Pneumonia	0.6333	0.5493	0.599	0.51	0.6333	0.62
Pneumothorax	0.7891	0.7425	0.7824	0.7516	0.7891	0.795
Average of precision	0.6078	0.5582	0.5597	0.5469	0.6078	0.6383

The average precision of eight anomalies shows the outperformance of the proposed model with a rate of 0.6383 compared to the other models (Wang et al. [43], AlexNet, GoogLeNet, VGGNet-16, and ResNet-50). In addition, the integration between deep learning, big data, and the social networks in the healthcare system was a challenge in this work because a new approach to using machine learning prediction was introduced.

To date, researchers have only focused on deep learning models, their features, targets, and precision. In this study, the deep learning model was integrated for the prediction of social network data. This integration can facilitate the use of the deep learning model for users of the social networking system.

The benefits of integrating deep learning into the healthcare system diagnosis are the prediction of diagnostics and diseases, as well as the shared deep learning knowledge (model) between users of social networks (patients, doctors, nurses). The deep learning model in this architecture can interact with social networks by providing comments on users' posts regarding the predicted diseases, which was a challenging paradigm in this proposal.

Moreover, a big data system for data storage and processing was proposed. The big data system stores healthcare data characterized by volume, variety, and velocity. Several deep learning models train its data in a central system. The proposed deep learning model trains three sets of data from distributed nodes. Thus, the distributed training approach is the most effective in practice.

Conclusion and future work

This paper presented a deep learning model based on big data for the healthcare system to predict social network data. The proposed architecture integrates three layers: deep learning, big data, and social networks.

Big data layer covers all issues related to large data, while deep learning layer is responsible for all deep learning algorithms. The social networks layer is used for social network users. This work also describes how to use a deep learning algorithm for a large volume of social network data.

Three CNN algorithms were introduced to validate the functionality of the proposed architecture. The CNN1 model was used to predict the type of image, while the CNN2 model was applied to determine whether the chest X-ray image was normal or abnormal. Finally, the CNN3 model was used to predict the type of diseases.

Users of social networks can interact with the proposed system to identify the classification of X-ray images. For future work, other issues related to the healthcare system will be examined to illustrate the benefits of the proposed architecture.

Ethical approval: The conducted research is not related to either human or animal use.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Conflict of interest: The authors declare that they have no conflict of interest.

References

[1] Schein R, Wilson K, Keelan JE. Literature review on effectiveness of the use of social media: a report for Peel Public Health. [Region of Peel], Peel Public Health, 2011.

[2] Expand your clinical knowledge, view and discuss real world clinical cases with healthcare professional from every specialty from Figure 1. Available at: https://figure1.com.
[3] Open biomedical initiative. Available at: http://www.openbiomedical.org.
[4] The National Library of Medicine presents MedPix. Available at: https://medpix.nlm.nih.gov/home.
[5] Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015;26:159–69.
[6] Achrekar H, Gandhe A, Lazarus R, Yu SH, Liu B. Predicting flu trends using twitter data. In 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) 2011:702–7. DOI: 10.1109/INFCOMW.2011.5928903.
[7] Hirose H, Wang L. Prediction of infectious disease spread using Twitter: a case of influenza. In 2012 Fifth International Symposium on Parallel Architectures, Algorithms and Programming 2012;100–5. DOI: 10.1109/PAAP.2012.23.
[8] Koh HC, Tan G. Data mining applications in healthcare. J Healthcare Inform Manag 2011;19:65.
[9] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Informatics 2008;77:81–97.
[10] Canlas RD. Data mining in healthcare: current applications and issues. School of Information Systems & Management, Carnegie Mellon University, Australia, 2009.
[11] Hosseinkhah F, Ashktorab H, Veen R, Owrang O. M. Mehdi. Challenges in data mining on medical databases. In Database Technologies: Concepts, Methodologies, Tools, and Applications 2009;4:1393–1404.
[12] Kumari M, Godara S. Comparative study of data mining classification methods in cardiovascular disease prediction, 2011.
[13] Dangare CS, Apte SS. Improved study of heart disease prediction system using data mining classification techniques. Int J Comp Appl 2012;47:4744–8.
[14] Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian J Comp Sci Eng 2011;2188–95.
[15] Khan MU, Choi JP, Shin H, Kim M. Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008:5148–51. DOI: 10.1109/IEMBS.2008.4650373.
[16] Chien C, Pottie GJ. A universal hybrid decision tree classifier design for human activity classification. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2012;1065–8. DOI: 10.1109/EMBC.2012.6346118.
[17] Moon SS, Kang SY, Jitpitaklert W, Kim SB. Decision tree models for characterizing smoking patterns of older adults. Exp Syst 2012;39:445–51.
[18] Chang CL, Chen CH. Applying decision tree and neural network to increase quality of dermatologic diagnosis. Expert Syst Appl 2009;36:4035–41.
[19] Fei SW. Diagnostic study on arrhythmia cord is based on particle swarm optimization-based support vector machine. Exp Syst Appl 2010;37:6748–52.
[20] Huang CL, Liao HC, Chen MC. Prediction model building and feature selection with support vector machines in breast cancer diagnosis. Exp Syst Appl 2008;34:578–87.
[21] Avci E. A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier. Exp Syst Appl 2009;36:10618–26.
[22] Abdi MJ, Giveki D. Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. Eng Appl Artif Intel 2013;26:603–8.
[23] Er O, Yumusak N, Temurtas F. Chest diseases diagnosis using artificial neural networks. Exp Syst Appl 2010;37:7648–55.
[24] Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Exp Syst Appl 2009;36:7675–80.
[25] Gunasundari S, Baskar S. Application of artificial neural network in identification of lung diseases. In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC) 2009:1441–4. DOI: 10.1109/NABIC.2009.5393702.
[26] Liu KF, Lu CF. BBN-based decision support for health risk analysis. In 2009 Fifth International Joint Conference on INC, IMS and IDC 2009;696–702.
[27] Gennings C, Ellis R, Ritter JK. Linking empirical estimates of body burden of environmental chemicals and wellness using NHANES data. Environ Int 2012;39:56–65.
[28] Agarwal S. Weighted support vector regression approach for remote healthcare monitoring. In 2011 International Conference on Recent Trends in Information Technology (ICRTIT) 2011:969–74. DOI: 10.1109/ICRTIT.2011.5972437.
[29] Tapia JJ, Morett E, Vallejo EE. A clustering genetic algorithm for genomic data mining. In Foundations of Computational Intelligence 2009;249–75.
[30] Escudero J, Zajicek JP, Ifeachor E. Early detection and characterization of Alzheimer's disease in clinical scenarios using bioprofile concepts and K-means. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2011;6470–3.
[31] Chipman H, Tibshirani R. Hybrid hierarchical clustering with applications to microarray data. Biostatistics. 2005;21:286–301.
[32] Chen TS, Tsai TH, Chen YT, Lin CC, Chen RC, Li SY, et al. A combined K-means and hierarchical clustering method for improving the clustering efficiency of microarray. In 2005 International Symposium on Intelligent Signal Processing and Communication Systems 2005;405–8.
[33] Belciug S. Patients length of stay grouping using the hierarchical clustering algorithm. Ann Univ Craiova-Math Comput Sci Ser 2009;36:79–84.
[34] Celebi ME, Aslandogan YA, Bergstresser PR. Mining biomedical images with density-based clustering. In International conference on information technology: coding and computing (ITCC'05)-volume II 2005;1:163–8.
[35] Ji Y, Ying H, Tran J, Dews P, Mansour A, Massanari RM. Mining infrequent causal associations in electronic health databases. In 2011 IEEE 11th International Conference on Data Mining Workshops 2011;421–8.
[36] Soni S, Vyas OP. Using associative classifiers for predictive analysis in health care data mining. Int J Comp Appl 2010;4:33–7.
[37] Widmer A, Schaer R, Markonis D, Müller H. Gesture interaction for content-based medical image retrieval. In Proceedings of International Conference on Retrieval 2014;503.

[38] Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TE, Bucholz R, et al. The Human Connectome Project: a data acquisition perspective. Neuroimage 2012;1:2222–31.
[39] Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. In Proceedings of the International Conference on Machine Learning, 2013.
[40] Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6:26094.
[41] Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. CoRR. arXiv preprint arXiv:1511.03677, 2015.
[42] Phan N, Dou D, Piniewski B, Kil D. Social restricted Boltzmann machine: human behavior prediction in health social networks. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015;424–31.
[43] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chest X-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on computer vision and pattern recognition 2017:2097–106.
[44] Chest X-ray dataset. Available at: https://ceb.nlm.nih.gov/repositories/tuberculosis-chest-X-ray-image-data-sets.
[45] LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, et al. Handwritten digit recognition with a back-propagation network. In Advanced Neural Information Processing Systems 1990:396–404.
[46] Python. Available at: https://python.com.
[47] Hadoop. Available at: http://hadoop.apache.org/.
[48] TensorFlow. Available at: https://www.tensorflow.org/install/.