Pneumonia Detection in Pediatric Chest X-Rays with Capsule Neural Networks

[1]Robert Langenderfer, [2]Ezzatollah Salari, [3]Jared Oluoch
[1,2]Dept. of Electrical Engineering and Computer Science, University of Toledo, Toledo, Ohio, USA
[3]Dept. of Engineering Technology, University of Toledo, Toledo, Ohio, USA

Abstract
Pneumonia is a common respiratory infection caused by bacteria, viruses, chemicals, or fungi. In 2019, more than 2.5 million people died from this disease, and is the single largest cause of death in children under the age of five globally. Diagnosis of pneumonia typically employs chest radiography, which is visually interpreted by highly trained radiologists. Given the cost, unavailability, and fallibility of radiologists, there has been significant interest in developing machine learning models to automate the diagnostics process. Recent research has focused on Deep Learning Neural Network (DNN) and Convolutional Neural Network (CNN) models to perform medical diagnostic classification. However, in this study we deployed a capsule based neural network for the detection of pneumonia in pediatric chest X-ray images. Where traditional CNN models discard significant image feature information due to thepooling layers, capsule networks preserve more information by utilizing vector outputs that encode the probability and pose for an observation. By preserving pose information, capsule networks preserve the spatial relationships between features and are immune translations, rotations, and scaling transformations of image data. This approach was evaluated on the publicly available pneumoniaMNIST radiological dataset. The proposed method achieved a verification accuracy of 98.3%, which exceeds the performance of models such as ResNet-18, ResNet-50, auto-sklearn, AutoKeras, and Google AutoML.

Keywords
Capsule Neural Network, Pneumonia, X-Ray, Pediatric

I. Introduction

Pneumonia is the most prevalent cause of mortality in young children globally [1]while in high-income countries it is one of the most common reasons for clinic attendance and hospitalization in this age group. Furthermore, pneumonia in children increases the risk of developing chronic pulmonary disorders in later adult life. While substantial advances in managing childhood pneumonia have been made, many issues remain, some of which are highlighted in this perspective. Multiple studies are required as many factors that influence outcomes, such as etiology, patient characteristics, and prevention strategies can vary between and within countries and regions. Also, outside of vaccine studies, most randomized controlled trials (RCTs. In 2019, more than 2.5 million people died from pneumonia, and is a huge burden on our healthcare systems as one of the top ten most expensive inpatient conditions to diagnose and treat. This disease is more prevalent in developing countries disadvantaged by a lack of proper medical facilities and environmental pollution. Pneumonia is also a major problem in developed countries, since pneumonia is the most common reason for hospitalization in the US, excluding childbirth [2].

Pneumonia is an inflammatory response to a respiratory infection which can be caused by bacteria, viruses, chemicals, or fungi. Pneumonia causes the lung alveoli to fill with fluids or cells, known as infiltrates, causing a reduction in diffusion capacity which in turn reduces blood oxygenation levels. The presence of infiltrates is detectable by X-rays and creates a discernable attenuation in radiological images due to the higher ratio of soft tissue to gas (increased density) relative to that of a healthy lung. Chest X-ray images are an important tool in the diagnosis of pneumonia [3].

Diagnosis typically requires a visual interpretation of the chest X-ray images by highly trained radiologists. The World Health Organization (WHO) estimates that there is a shortage of 4.3 million health professionals globally. Developing nations, rural, and remote regions suffer the most from this issue [4] the frequency and clinical impact of errors in the anatomic pathology diagnosis of cancer have been poorly characterized to date. METHODS. The authors examined errors in patients who underwent anatomic pathology tests to determine the presence or absence of cancer or precancerous lesions in four hospitals. They analyzed 1 year of retrospective errors detected through a standardized cytologic-histologic correlation process (in which patient same-site cytologic and histologic specimens were compared. Misdiagnosis is also a serious issue as, in one study, 72 percent of patients were misdiagnosed with pneumonia upon readmission to the same hospital [5]. These reasons provide ample motivation for the implementation of automated image processing techniques for pneumonia classification.

Machine learning techniques have been used extensively in the classification of a variety of illnesses, utilizing data from a range of medical diagnostic tools [6]. Deep learning algorithms can be used to detect and diagnose pneumonia using only X-ray images, and in the process saves both money and time. Medical professionals can also benefit from this approach by efficiently identifying highly critical patients and reducing misdiagnoses. Recent research has focused mostly on Deep Learning Neural Network (DNN) and Convolutional Neural Network (CNN) models to perform this diagnostic classification.

Traditional CNN models are composed of numerous 2-dimensional layers which propagate weighted scalar values through the network. There are two basic types of layers in a CNN: convolutional and pooling. Convolutional layers utilize 'kernels': square scalar arrays that are iterated over an image, which is a technique utilized in traditional 2D image processing algorithms [7]. These scalar parameter values are learned by the network through error backpropagation. The impact of these kernels is to emphasize various characteristics of the image, such as vertical, horizontal lines, corners, etc. Note again that the network learns the kernel parameters, so the network 'decides' the optimal values. Following the convolutional layers are the pooling layers. The pooling layer serves to compress the images by reducing each kernel output to a single scalar value[8]. Here lies the source of one of the most

serious criticisms of CNN's, the data reduction in the pooling layers discard a significant amount of information about the image. The impact of this data loss is known as the 'Picasso problem', meaning that the relationships between the features lose their spatial relativity, thus even an obviously scrambled image can still be considered valid by a CNN [9]. Also, CNNs are sensitive to transformations such as 2D and 3D rotations, scale, translation, and skew. To overcome this limitation, CNNs are typically trained on augmented data, whereby the images in the original dataset are subject to a range of thetransformations listed above. While effective, this augmentation dramatically increases the effective training dataset size, which in turn increases the training time [10].

To overcome the limitations of CNN models, we propose to implement the Capsule Neural Network model. This model is intended to improve the classification performance and will be tested on X-Ray image dataset. In section II the Capsule networkwill be introduced followed by the implementation and dataset details in section III, then the results in section IV, and the conclusion in section V.

II. Proposed Method

A. Capsule Neural Networks

The Capsule network is a relatively new deep learning neural technique created by Geoffrey Hinton that operates in a significantly different manner than conventional CNN methods. Capsule networks differfrom CNNs in three primary ways:

1. The network layers use weighted vectors, not scalars
2. The capsule layers route by agreement, and
3. The network performs inverse graphics [11].

There are two basic parts to the network: the encoder and the decoder.

B. The Encoder

The training images are first processed by the initial two layers of the encoder. These initial layers consist of two convolutional layers identical to those composing a CNN. However, that's where the similarity to CNNs end. Capsule network layers utilize vectors instead of scalar values which are weighted according to the following:

$$\hat{u}_{ji} = W_{ij} u_i$$

The output of capsule i is the vector u_i which is multiplied by the scalar weight matrix W_{ij} and produces the vector $\hat{u}_{ji}$ output for the next level capsule j. These vectors encode the probabilities of a specific object through their lengths and the vectors directions encode the state of the detected objects, such as rotation. Now capsule networks use a new connection method known as 'routing by agreement.' The capsules in each layer attempt to predict the output of the next layer [12]. The better the accuracy of the prediction, or correlation, the more the capsules are connected through the 'coupling coefficient' c_{ij}. The output vector û_ji is multiplied by this coupling coefficient and is subject to the following summation formula

$$s_j = \sum_i c_{ij}\, \hat{u}_{ji}$$

where s_j is the input of capsule j into the next layer. The value of c_{ij} above is based on the formula

$$softmax(b_{ij}) = c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

where b_{ij} is the probability of the connection between capsule i and capsule j. Note that b_{ij} starts out at zero, so there is initially no effective connection between capsule layers. This softmax function enhances the largest values and suppresses values which are significantly below the maximum value, while scaling the vector such that the outputs add up to 1.

Traditional neural networks subject neuron output into a non-linear activation function such as Relu or sigmoid, but those functions can't be used on vectors, so the Capsule network implements a squashing function instead:

$$squash(s_j) = v_j = \frac{||s_j||^2}{1 + ||s_j||^2} \frac{s_j}{||s_j||}$$

where v_j is the final output of capsule j. This formula normalizes the vector components to the range of 0 to 1[13]. This function produces a curve that resembles the upper half of the sigmoid function, when viewed from 2 dimensions. The probabilities b_{ij} from the softmax (b_{ij}) function must be updated using the dot product of v_j and $\hat{u}_{ji}$.

$$b_{ij} = b_{ij} + v_j \cdot \hat{u}_{ji}$$

This equation is the key to routing by agreement because the greater the alignment of the two vectors the more the coupling coefficient is increased since it is based on b_{ij}. [14]

The encoder portion of the network uses the following margin loss function

$$L_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k)\max(0, ||v_k|| - m^-)^2$$

The hyperparameters m^+, m^-, and λ are constants assigned to the values 0.1, 0.9, and 0.5 respectively, and T_k = 1 only when a category of class k is present. The loss function is applied to each of the capsules, and the total loss is simply $\sum_k L_k$. [15]

The number of capsules in the last layer of the encoder correspond to the number of categories in the dataset. For example, a binary classification will have two last layer capsules. This last layer feeds into the encoder portion of the network.

Table 1: Dynamic Routing Function Pseudo-Code

Dynamic routing function
Routing($\hat{u}_{ji}$, r, l)
foreach capsule i in layer l
foreach capsule j in layer (l + 1)
$b_{ij} \leftarrow 0$
foreach r
foreach capsule i in layer l
$c_i \leftarrow softmax(b_{ij})$
foreach capsule j in layer (l + 1)
$s_j \leftarrow \sum_i c_{ij}\, \hat{u}_{ji}$
foreach capsule j in layer (l + 1)
$v_j \leftarrow squash(s_j)$
foreach capsule i in layer l
foreach capsule j in layer (l + 1)
$b_{ij} \leftarrow b_{ij} + v_j \cdot \hat{u}_{ji}$
return v_j

C. The Decoder

The decoder is composed of fully connected network layers whose final output matches the size of the input layer. The encoder recreates the input images and trains the network using backpropagation the use of a loss function that is simply the N-dimensional Euclidian distance between the decoded image and the input image. This image generation technique is referred to as inverse graphics. A diagram of the network is depicted in Fig. 1.

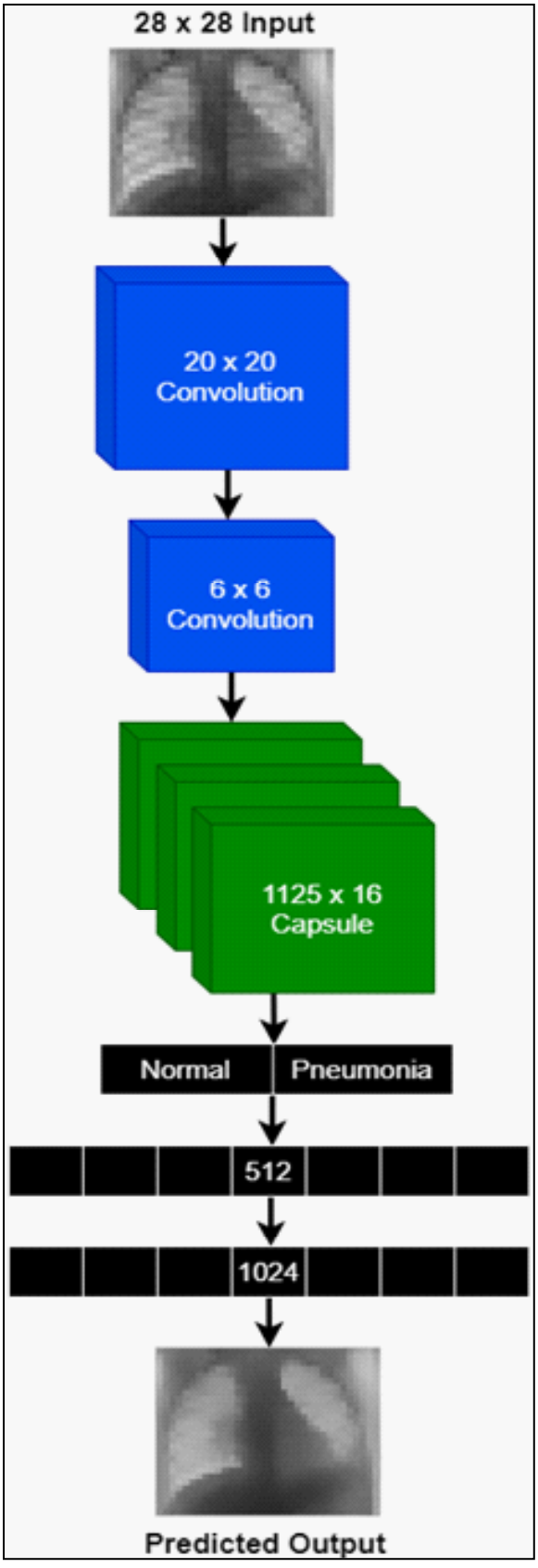

Fig. 1: Capsule Network Diagram

III. Implementation

A. Pneumonia Dataset

The capsule network described herein was trained and validated with the open source pneumoniaMNIST dataset, which is a subset of the MedMNIST v2 data [16]. This grayscale dataset is based on a 5,856 pediatric chest X-Ray images which have a range of sizes: (384−2,916) × (127− 2,713). An example of full resolution pneumonia and normal images are shown in Figure 2. These source images have been center cropped and resized to 28×28, with 256 levels of grayscale. Figure 3 displays the same two high resolution images from Fig. 2 but resized to 28x28 pixel resolution. Note that this resizing causes adramatic reduction in pixel data, blurring fine anatomical details a radiologist would easily view in the high-resolution image.

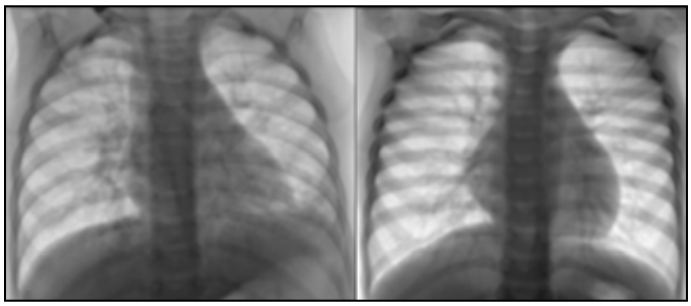

Fig. 2: Full Resolution X-Ray Images Pneumonia (left) Normal (right)

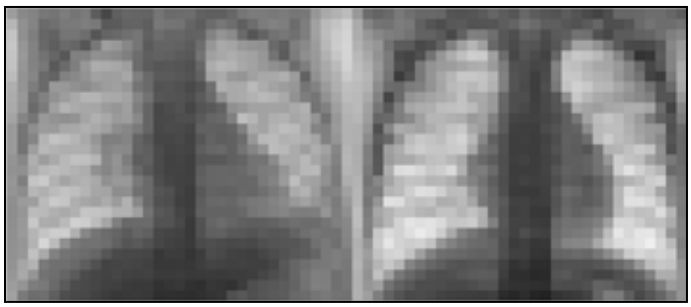

Fig. 3: Resized 28 x 28 X-Ray Images Pneumonia (left) Normal (right)

B. Programming and Training

The Capsule network was implemented in Python using Tensorflow, Keras, and Sklearn, which are all open-source libraries. The network was trained on Kaggle using GPU acceleration.

The dataset was divided into 5,332 training images and 524 test images. The original, full-size images were classified by a team of radiologists into two categories: normal or pneumonia. There are 1,835 normal and 3,497 pneumonia examples in the training set and 135 normal and 389 pneumonia examples in the test set. There was no data augmentation performed on this dataset.After training, the decoder section of the network outputs similarly sized recreations of the input images. Some examples of these predicted outputs and their corresponding input images are shown in Fig. 4,

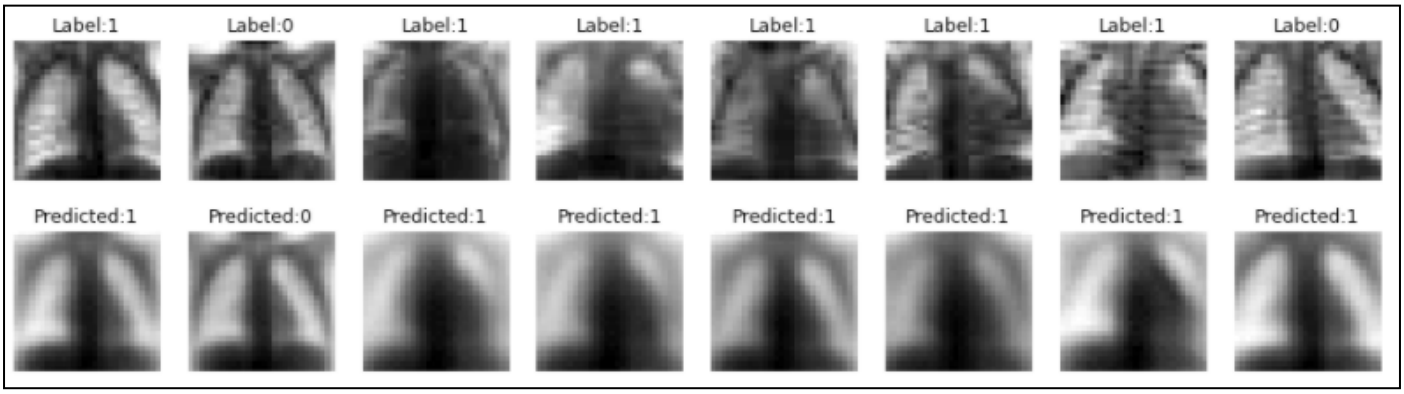

Fig. 4: Input Images with Predicted Images

IV. Results

The performance of the network was evaluated using the accuracy, sensitivity, and area under the curve (AUC) metrics. Accuracy and sensitivity are calculated as follows:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$sensitivity = \frac{TP}{TP + FN} \cdot 100\%$$

Where TP is the number of true positives, TN the number of true negatives, and FN is the number of false negatives, and FP the false positives. These values and their percentages are listed in the confusion matrix of Fig. 5.

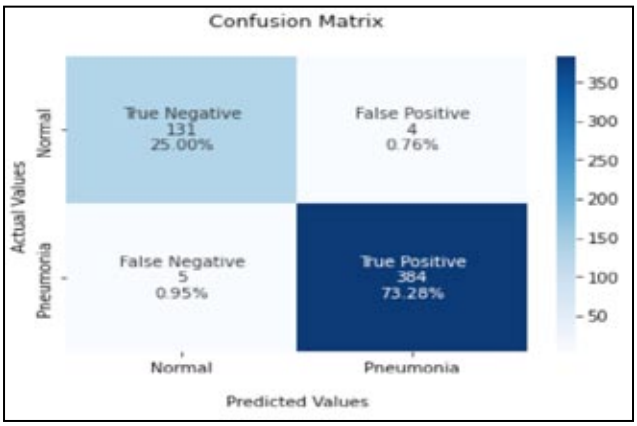

Fig. 5: Confusion Matrix

The AUC value is calculated by finding the Receiver Operating Curve (ROC) which is area under the plot of the True Positive Rate (TPR) against the False Positive Rate (FPR) at a range of threshold settings. The ROC curve is depicted in Fig. 6.

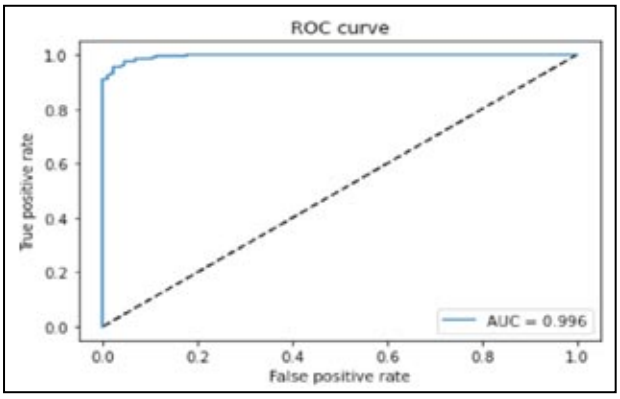

Fig. 6: The ROC curve is depicted

The performance metric results are listed in Table 2. The AUC and accuracy numbers are compared to several of the popular CNN models, such as ResNet and Google AutoML Vision, as shown in Table 3 [16]a large-scale MNIST-like dataset collection of\nstandardized biomedical images, including 12 datasets for 2D and 6 datasets for\n3D. All images are pre-processed into a small size of 28x28 (2D. The capsule network outperformed the top model, Google AutoML Vision, by 3.7%.

Table 2: Performance Metrics

Metric		
AUC	**Accuracy**	**Sensitivity**
.996	98.3%	98.7%

Table 3: Capsule Network Performance Comparison

Performance Comparison		
Network	**AUC**	**Accuracy**
ResNet-18 (28)	0.944	0.854
ResNet-18 (224)	0.956	0.864
ResNet-50 (28)	0.948	0.854
ResNet-50 (224)	0.962	0.884
auto-sklearn	0.942	0.855
AutoKeras	0.947	0.878
Google AutoML Vision	0.991	0.946
Capsule Network	**0.996**	**0.983**

V. Conclusion

Pneumonia is the most prevalent cause of mortality in young children globally. For the accurate diagnosis and treatment of pneumonia, knowledge of the state of a patient's lungs are vital. Radiological imaging is an effective method of revealing lung functionality, and the state of disease can be successfully diagnosed. The open source pneumoniaMNIST database, which is composed of 5,856 28x28 grayscale pediatric radiology images, was utilized for training and testing data. By implementing machine learning classification using Capsule neural networks on these images, machine diagnostic efficacy that rivals that of human experts was demonstrated. Despite the low resolution of the dataset, the method described was able to achieve a verification accuracy of 98.3%. These results are a significant improvement in accuracy over other state of the art methods and did not require data augmentation typical of CNN models.

References

[1] A. B. Chang, M. H. Ooi, D. Perera, and K. Grimwood, "Improving the diagnosis, management, and outcomes of children with pneumonia: Where are the gaps?," Front. Pediatr., Vol. 1, No. OCT, pp. 29, Oct. 2013.

[2] C. L. Fischer Walker et al., "Global burden of childhood pneumonia and diarrhoea," Lancet, Vol. 381, No. 9875, pp. 1405–1416, 2013.

[3] A. T. Society, "Top 20 Pneumonia Facts — 201 9," 2016.

[4] S. S. Raab et al., "Clinical impact and frequency of anatomic pathology errors in cancer diagnoses," Cancer, Vol. 104, No. 10. pp. 2205–2213, 2005

[5] H. F. H. System,"Pneumonia often misdiagnosed on patient readmissions, studies find -- ScienceDaily." [Online] Available: https://www.sciencedaily.com/releases/2010/10/101022123749.htm

[6] G. Litjens, "A survey on deep learning in medical image analysis," Med. Image Anal., Vol. 42, pp. 60–88.

[7] L. G. Shapiro, G. C. Stockman,"Computer vision," pp. 580, 2001.

[8] A. Krizhevsky, I. Sutskever, G. E. Hinton,"ImageNet classification with deep convolutional neural networks," Commun. ACM, Vol. 60, No. 6, pp. 84–90, 2017.

[9] V. Gliozzi, G. L. Pozzato, A. Valese,"Combining neural and symbolic approaches to solve the Picasso problem: A first step," Displays, Vol. 74, pp. 102203, 2022.

[10] C. Shorten, T. M. Khoshgoftaar,"A survey on Image Data Augmentation for Deep Learning," J. Big Data, Vol. 6, No. 1, pp. 1–48, 2019.

[11] G. E. Hinton, A. Krizhevsky, S. D. Wang,"Transforming Auto-encoders."

[12] "Introducing capsule networks – O'Reilly." [Online] Available: https://www.oreilly.com/content/introducing-capsule-networks/

[13] S. Sabour, N. Frosst, G. E. Hinton,"Dynamic Routing Between Capsules," Adv. Neural Inf. Process. Syst., pp. 3857–3867, 2017.

[14] E. Xi, S. Bing, Y. Jin, "Capsule Network Performance on Complex Data," 2017.

[15] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," Pattern Recognition Letters, vol. 138. Elsevier B.V., pp. 638–643, 2020.

[16] J. Yang et al., "MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification," 2021.

Robert Langenderfer received his B.S. degree in Computer Science and Engineering and M.S. degree in Engineering Science and is currently a PhD candidate in Computer Science and Engineering at the University of Toledo, Toledo, Ohio, USA. He is an assistant lecturer of Computer Science and Engineering Technology at the University of Toledo.He has many years of industry experience developing machine learning, gaming, web, and E-commerce applications for numerous Fortune 500 companies and major Universities. His research interests include machine learning algorithms, automated medical diagnostics, neural networks applied to image processing, and intelligent control algorithms.

Dr. Ezzatollah Salari received his M.S. and Ph.D. degrees in Electrical Engineering from Wayne State University in 1978 and 1982, respectively. He is at present a professor in the Department of Electrical Engineering and Computer Science at the University of Toledo where he is involved in teaching and research in the areas of image analysis and computer vision, data compression for multimedia communication, neural networks, and signal processing. He has contributed extensively to several areas of image processing and neural networks including motion analysis, image representation, image and video compression, and applications of neural networks in image processing. He served as Graduate Director of the EECS Department at UT from 2000 to 2005. Dr. Salari also worked as a research fellow at NASA Langley, Goddard, and Lewis (Glenn) research centers during the summers of 1987, 1988, 1990 and 1991 on various projects.

Jared Oluoch is an Associate Professor of Computer Science and Engineering Technology (CSET) at the University of Toledo, OH. He received his PhD in Computer Science and Informatics from Oakland University in Rochester, Michigan in 2015. His research interests span five broad areas: 1) Object-oriented programming for machine learning algorithms, 2) Reputation management and security of Connected and Autonomous Vehicles (CAVs), 3) Cyber Security, 4) Localization algorithms for Wireless Sensor Networks (WSNs), and 5) Engineering education. He has secured external research funding of over $2,000,000 dollars from NSF as the Principal Investigator. He has published in peer-reviewed conferences and journals as a primary author or co-author in his areas of research. He has advised and currently advises several masters and doctoral students in Computer Science. He has been nominated for five consecutive years for the University of Toledo Outstanding Advisor award. He has served on several NSF panels as a reviewer. He is a senior member of IEEE and a Program Evaluator for ABET Engineering Technology Accreditation Commission (ETAC).

.