Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(6), 3235-3242 | ISSN:2251-6727

Ovarian Cancer Subtypes Reimagined: Leveraging CNNs for Advanced Classification and Outlier Detection

Dr. R. Ramya[1], Karthik Reddy Duddukunta[2], Sai Vikas Addanki[3], Vishnu Sai Bezawada[4]

[1,2,3,4]Dept .Of Computer Science And Engineering, SRMIST , Ramapuram, Chennai, India

(Received: 07 October 2023 Revised: 12 November Accepted: 06 December)

KEYWORDS

Ovarian cancer, CNN, Deep Learning, Classification, Detection.

ABSTRACT:

Ovarian cancer remains an aggressive challenge in the realm of cancer research and treatment due to its diverse subtypes, each characterized by distinct genetic and molecular features. The UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN) applies advanced deep learning techniques, particularly convolutional neural networks (CNNs), to address the complexities of ovarian cancer subtypes like CC, EC, HGSC, LGSC, and MC classification and outlier detection. A robust CNN-based model capable of accurately classifying different subtypes of ovarian cancer. Traditional methods often face challenges in handling the complicated patterns and subtle differences in medical imaging data. CNNs excel in image-based tasks by automatically learning hierarchical representations, making them ideal for the nuanced classification required in cancer subtype analysis. Subtype classification: the UBC-OCEAN incorporates outlier detection mechanisms to identify anomalies within the dataset. Outliers may represent rare or previously unidentified subtypes or instances of the disease that exhibit unique characteristics. Integration of advanced deep learning techniques in cancer research signifies a paradigm shift towards more sophisticated and efficient approaches in the quest for improved diagnostic accuracy and patient outcomes. The UBC-OCEAN represents a crucial step forward in harnessing the power of CNNs for the intricate task of ovarian cancer subtype classification and outlier detection.

1. INTRODUCTION

Ovarian cancer is a prevalent and deadly gynaecologic cancer with a high rate of late-stage diagnosis, resulting in low survival rates. This study explores the use of deep learning, specifically the Fare-CNN model, for the early detection and classification of ovarian cancer. The research involves a comprehensive literature review, discusses the limitations of convolutional neural networks with region proposal networks. The methodology includes data augmentation, segmentation, and classification, demonstrating the potential of deep learning in improving ovarian cancer diagnosis. The study highlights advancements in medical imaging and presents a comparative analysis of proposed and existing techniques [1].

This paper introduces an efficient Convolutional Neural Network (EFFI-CNN) designed for the detection of lung cancer using CT scan images from LIDC-IDRI and Mendeley datasets. With seven specialized layers, including Convolution, Max-Pool, Fully Connected, and Soft-Max, EFFI-CNN aims to enhance accuracy and minimize detection time. The proposed architecture systematically processes input images, leveraging a unique combination of layers. EFFI-CNN's design is motivated by the critical need for early lung cancer detection, addressing a significant health concern. This study builds upon advancements in deep learning, emphasizing the potential of CNNs in medical image analysis and showcasing promising results in lung cancer detection [2]. The paper emphasizes the need for sustainable ultrasound standards to reduce maternal and perinatal mortality rates. It introduces a machine learning approach, including Logistic Regression Classifier (LRC) and Convolutional Neural Networks (CNNs), for obstetric ultrasound analysis. The Internet of Medical Things (IoMT) is utilized for obstetric tumor segmentation. Experimental results demonstrate the effectiveness of LRC based on CNN in predicting ultrasound outputs, contributing to improved maternal and perinatal outcomes [3].

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

Ovarian cancer, ranking as the fifth most prevalent cancer in women, necessitates accurate detection and staging for effective treatment. This research proposes a robust methodology employing a Scale-Invariant Feature Transformation (SIFT) algorithm for feature extraction. A Genetic Algorithm optimizes these features based on a fitness function, considering individual features, total features, and classification error rate. Convolutional Neural Network (CNN) classification achieves an impressive 98.8% accuracy, outperforming Support Vector Machine (SVM). The system excels in sensitivity, specificity, and accuracy, offering a promising solution for ovarian cancer detection and staging [4]. Ovarian cancer, a rising concern in women's health, demands effective detection methods for early intervention. This review explores the significance of ovaries, their role in cancer development, and the escalating rates of ovarian cancer. The focus lies on leveraging Convolutional Neural Networks (CNN) for accurate tumour classification.

A comparative analysis with machine learning algorithms like K-Nearest Neighbour and Support Vector Machine underscores the superiority of Deep Learning, particularly CNN, in ovarian cancer detection. The paper provides insights into the challenges, symptoms, and stages of ovarian cancer, emphasizing the pivotal role of advanced technologies in combating this prevalent health issue [5]. This paper addresses the challenge of colorectal cancer cell detection in medical images. It proposes an enhanced Faster R-CNN algorithm with multi-scale detection and a multi-loss function to overcome issues related to complex backgrounds and small object cells. The multi-scale detection method utilizes both low-level and high-level features for better object representation, while the complex loss function minimizes intra-class distance, improving discriminative feature learning. Experimental results demonstrate a 2.4% accuracy improvement compared to Faster R-CNN. The proposed MC Faster R-CNN algorithm offers significant advancements in colorectal cancer cell detection, showcasing its high application value [6]. Ovarian cancer, with its diverse subtypes characterized by unique genetic and molecular features, presents a formidable challenge in cancer research and treatment. Addressing this complexity, the UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN) employs advanced deep learning techniques, specifically convolutional neural networks (CNNs), to enhance the classification and outlier detection of subtypes such as CC, EC, HGSC, LGSC, and MC.

Traditional methods in cancer subtype analysis struggle with the intricate patterns and subtle variations in medical imaging data. CNNs prove highly effective in this context by autonomously learning hierarchical representations, making them well-suited for nuanced cancer subtype classification. UBC-OCEAN incorporates outlier detection mechanisms to identify anomalies within the dataset, potentially revealing rare or previously unidentified subtypes. The proposed CNN model stands out for its ability to handle large datasets efficiently. In contrast to artificial neural networks (ANN), which are computationally intensive and treat distant pixels similarly, CNNs leverage filter operations, particularly beneficial for grid-like matrix datasets such as RGB images. The CNN architecture comprises input, hidden, and output layers.

The convolutional layer, a cornerstone of CNNs, employs learnable filters or kernels to extract features from input images, facilitating nuanced pattern recognition. Activation functions, such as Rectified Linear Unit (ReLU), introduce non-linearities critical for learning complex relationships within the data. The subsequent pooling layer reduces spatial dimensions, enhancing computational efficiency. The UBC-OCEAN CNN model, with its emphasis on CNN architecture, advanced pooling mechanisms, and fully connected layers, represents a significant leap in leveraging deep learning for accurate ovarian cancer subtype classification and outlier detection. The integration of such sophisticated techniques marks a paradigm shift toward more efficient and precise approaches in cancer research, promising improved diagnostic accuracy and patient outcomes.

2. LITERATURE SURVEY

ovarian cancer detection, emphasizing the limitations and challenges faced by conventional approaches. It critically reviews studies on image-based diagnostic techniques, discussing the role of machine learning and deep learning in enhancing early detection accuracy. The survey encompasses a range of relevant works, providing a comprehensive understanding of the current landscape and paving the way for the proposed Fare-CNN model's potential contributions to advancing ovarian cancer diagnosis [1]. This method proposes EFFI-CNN, an efficient Convolutional Neural Network, for lung cancer detection using CT scan images from LIDC-IDRI and Mendeley datasets. The seven-layer architecture combines Convolution, Max-Pool, Fully Connected, and Soft-Max layers to enhance accuracy and minimize detection time. Inspired by the growing significance of deep learning in medical image analysis, EFFI-CNN

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

aims to outperform existing literature. The work underscores the urgent need for early lung cancer detection, given the alarming statistics provided by the National Institutes of Health. Utilizing advanced neural networks, EFFI-CNN showcases promising results, contributing to the evolving landscape of medical image analysis [2].

This research provides a convolutional neural network model for obstetric imaging ultrasonography that is based on machine learning and may detect ovarian cancers during pregnancy. With a set of tiny filters, typically of size 3–3, in a tensor S (i, j), the model transforms activations from earlier layers into convolutional layers. This significantly lowers the number of weights that must be learned because characteristics in one area of the image might be present in other areas. This CNN aims to provide a function that can categorize an input patch into four groups [3].

The main objective of this work is to analyze 512*512 picture data of ovarian cancer utilizing the fuzzy c-means clustering algorithm, edge detection, and filtration techniques. The image acquisition, preprocessing, edge detection, segmentation, feature extraction, optimal features, and classification steps are all included in the methodology. Using cunning edge detection techniques, the image is chosen, pre-processed, and edge detected. The SIFT algorithm, which offers an object's "feature" description, is used to accomplish the feature extraction technique. The feature extraction matrix is reduced by the genetic method, and a fitness function is utilized to optimize the dataset. Convolutional neural networks are used for the classification, and performance metrics are assessed using the new methodology [4].

This review paper suggests a deep learning classifier, Convolutional Neural Network (CNN), along with a multi-network feature extraction model, DenseNet-201, for the classification of ovarian cancer from medical histopathology pictures. The model makes use of 24-bit TrueColor color space H&E-stained pictures of ovarian slides. A large training set is necessary to prevent overfitting, and the size of the data test determines how effective the Deep Learning approach is. The photos are classified using the CNN classifier that has been trained [5].

A region-based object detection network. Our approach enhances feature extraction by leveraging convolutional layers and pooling layers to extract information at different scales. Multiple feature maps obtained through the region proposal networks (RPN) contribute to the region proposal features, which are subsequently processed by the RoI pooling layer. For effective object detection, we introduce a novel loss function that considers the complex background of colorectal cancer cells. This loss function minimizes intra-class distances, ensuring discriminative features capture both background complexities and inter-class differences. The approach demonstrates superior learning capabilities for accurate cell classification and feature extraction [6]. Existing model explores diverse methods for disease classification using deep learning. Automated techniques, such as CNN with graph-cuts for lung CT scans and 3D-GLCM-CNN for polyp recognition, showcase advancements. Grid Search-based Hyper-Parameter Tuning and Evolutionary multi-objective optimization tools offer solutions in diverse domains. Multi-task multi-scale deep learner systems and integrated approaches demonstrate progress, emphasizing the need for effective algorithms. This survey provides insights into current methodologies, setting the stage for the proposed DSSGL-EUNet-MSDCNN model in ovarian cancer diagnosis [7]. The proposed UBC-OCEAN signifies a paradigm shift in cancer research, demonstrating the potential of CNNs for improved diagnostic accuracy and patient outcomes. Literature emphasizes CNNs' versatility in computer vision, medical imaging, and deep learning advancements, making them a pivotal tool in diverse research domains.

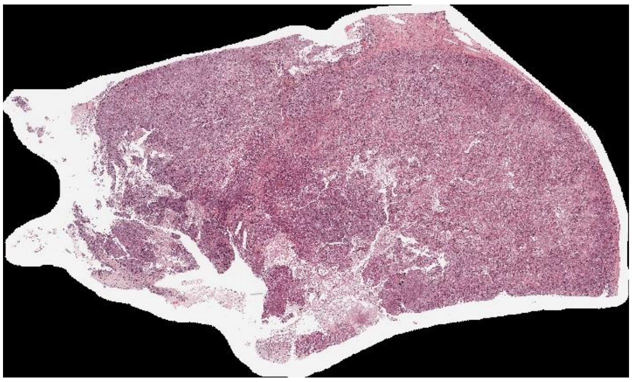
3. PROPOSED WORK

Our proposed CNN model is used to handle large amounts of data. ANN is also used to handle complex datasets, but ANN takes too much computation time, treats local pixels as if they were a part of the image, and is sensitive to the location of an object in the image. In CNN's model, the filter option, like RGB image, is nothing but a matrix of pixel values with three planes.
An extended neural network, or CNN, is used to extract features from a matrix dataset that resembles a grid. Multiple layers make up CNN, including the input layer, pooling layer, convolutional layer, and fully connected layers.

Input layer: It receives input from external sources. The image consists of 16 widths and 16 heights, encompassing three RGB channels or layers, and then it will hold the raw pixel values of the image.
The convolution layer consists of a set of learnable filters having the same widths and heights as the input volume; these are known as kernels for the input image. It computes the output of those layers thar are associated

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

with input's local region such that each layer will calculate a dot product.


For sample Fig. 1 is taken for convolutional layer operation

$$(n \times n) \times (f \times f) = (n - f + 1) \times (n - f + 1)$$

Fig.1 has width=3000 and height=1790

$$= 2998 \times 1778 \; image\ size\ after\ convolutional\ operation$$

n= size of image
f= filter

In this convolutional layer, stride is applied. Stride refers to the step size or the interval at which the convolutional filter (also known as the kernel or window) is moved across the input data. The stride in CNNs is a design decision that balances spatial information preservation, computational efficiency, and the trade-off between model complexity and representational power. The appropriate stride value depends on the specific requirements of the task and the characteristics of the data.

$[n - f \div s + 1]$ s=stride

CNNs use activation functions to add non-linearities to the model and enable the network to recognize complex patterns and relationships in the input. The network would simply be a linear model without activation functions, unable to represent the complex features found in images and other non-linear data. Numerous activation functions exist, such as leaky ReLU, sigmoid, hyperbolic tangent (Tanh), soft max, and Rectified Linear Unit (ReLU); however, ReLU is included in this model since it is the most often used activation function in CNNs. Positive numbers continue to go through unaltered, while all negative values in the input are replaced with zero. ReLU reduces the loss of gradient issue and aids in quicker training convergence.

$$Function: f(x) = \max(0, x)$$

Next, pooling is introduced to the CNN model. This layer is utilized to conduct a downsampling operation along the spatial dimensions, which aids in reducing the input volume's spatial dimensions by lowering the network's computational cost and parameter count. The max pooling and average pooling layers are the two most used varieties.

1. Max Pooling: Max pooling is a down sampling operation that extracts the most important information from a local region; the maximum value is retained, discarding the less relevant information. It will depend on stride, and it will take the maximum element.
2. Average Pooling: It is similar to Max Pooling, but instead of taking the maximum value from the local region, it computes the average.

However, only Max Pooling is employed in these CNN models due to the introduction of form translation invariance, which lessens the network's sensitivity to minute changes in position. The next important step is to flatten the multi-dimensional feature maps that the convolutional and pooling layers create into a one-dimensional vector. The network is able to provide predictions based on the high-level features that are retrieved from the input data by using this flattened vector as the input for fully connected layers.

The process of flattening entails organizing every value from the feature maps into a single, continuous vector.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

Then, from the high-level data that convolutional and pooling layers have collected, fully connected layers—also referred to as dense layers—play a critical role in learning intricate patterns and correlations. These layers enable the network to identify global patterns and provide predictions by connecting all of the neurons in one layer to all of the neurons in the subsequent layer. High-level representations of the input data are contained in this flattened vector.

Every neuron in a completely connected layer is linked to every other neuron in the layer that came before it. The parameters that the network learns throughout the training phase are the weights and biases connected to these connections. Depending on the particular task, the output layer of the completely connected layers has a different number of neurons. A sigmoid activation function for a single neuron is commonly employed for binary classification. The number of neurons in a multi-class classification system is equal to the number of classes, and a soft max activation is frequently used. Dimensionality reduction is achieved by fully connected layers, which compile the features that were retrieved from the convolutional and pooling layers into a format that can be used for prediction.

These layers' weights aid in the learning of intricate hierarchical representations. Every neuron in the layer above is coupled to every other neuron in a completely connected layer. During training, the network learns these parameters, which are the weights and biases connected to these connections. The particular job determines how many neurons are in the output layer of the completely linked layers. A single neuron with a sigmoid activation function is frequently utilized for binary classification. A soft max activation is frequently used in multi-class classification, where the number of neurons is equal to the number of classes.
Fully connected layers perform dimensionality reduction by summarizing the extracted features from the convolutional and pooling layers into a format suitable for making predictions. The weights in these layers help in learning complex hierarchical representations.

In the initial phase of the architecture, the process begins with obtaining a dataset from Kaggle, serving as the input for subsequent steps. Following this, data preprocessing and exploratory data analysis are conducted, involving the removal of null values and outliers to enhance the accuracy of data collection. The third step involves encoding the data and training the images, with a subsequent application of a Convolutional Neural Network (CNN) model for cancer subtype classification.

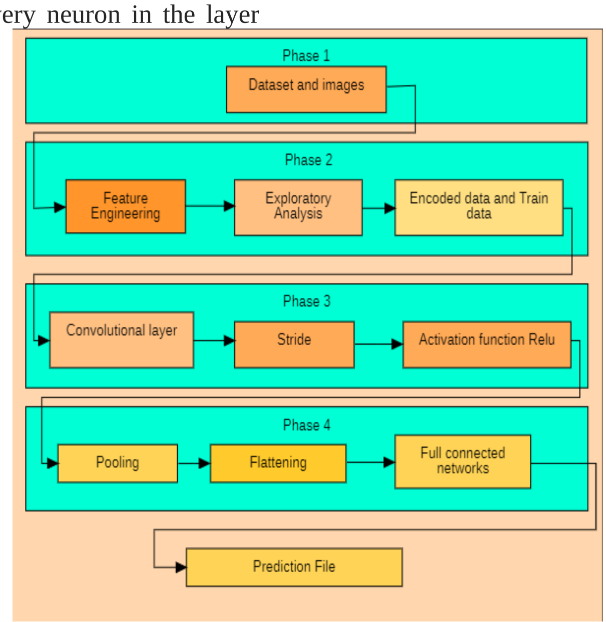


Fig 2: Architecture diagram for proposed work

In the CNN model, the first convolutional layer performs operations such as feature extraction on the input images. Stride and activation functions are then applied to fine-tune the convolutional layer's output. Subsequently, pooling and flattening operations are employed to prepare the data for integration into a fully connected network. This comprehensive approach leverages convolutional layers, activation functions, and pooling to

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

extract relevant features from the images, facilitating the effective classification of cancer subtypes. The orchestrated sequence of data preprocessing, model training, and CNN application collectively contributes to a robust and accurate cancer subtype classification system.

4. RESULTS AND DISCUSSION

There are 539 items in the dataset, and five different characteristics are included: label, image width, image height, image id, and tma. provide insights into the evolution of the model during training. The training loss and accuracy metrics indicate how well the model is learning from the training data, while the validation loss and accuracy metrics give an indication of the model's generalization to unseen data. In this case, it's notable that the validation accuracy remains relatively constant at 0.13, which might indicate a need for further model adjustments or investigation into potential issues such as overfitting. Fig. 2 shows the accuracy and loss of both training and validation of images and data.

```
Epoch 1/20
25/25 [==============================] - 3s 47ms/step - loss: 2.7050 - accur
acy: 0.0775 - val_loss: 2.3023 - val_accuracy: 0.0950
Epoch 2/20
 1/25 [>.............................] - ETA: 0s - loss: 2.3151 - accuracy:
0.0625

C:\Users\karth\anaconda3\envs\ML\Lib\site-packages\keras\src\engine\trainin
g.py:3079: UserWarning:

You are saving your model as an HDF5 file via `model.save()`. This file form
at is considered legacy. We recommend using instead the native Keras format,
e.g. `model.save('my_model.keras')`.


25/25 [==============================] - 1s 27ms/step - loss: 2.2996 - accur
acy: 0.1287 - val_loss: 2.6197 - val_accuracy: 0.1100
Epoch 3/20
25/25 [==============================] - 1s 29ms/step - loss: 2.2957 - accur
acy: 0.1138 - val_loss: 3.8203 - val_accuracy: 0.1100
Epoch 4/20
25/25 [==============================] - 1s 30ms/step - loss: 2.2833 - accur
acy: 0.1112 - val_loss: 5.1051 - val_accuracy: 0.1100
```

Fig 3: Accuracy and loss of train and validation data

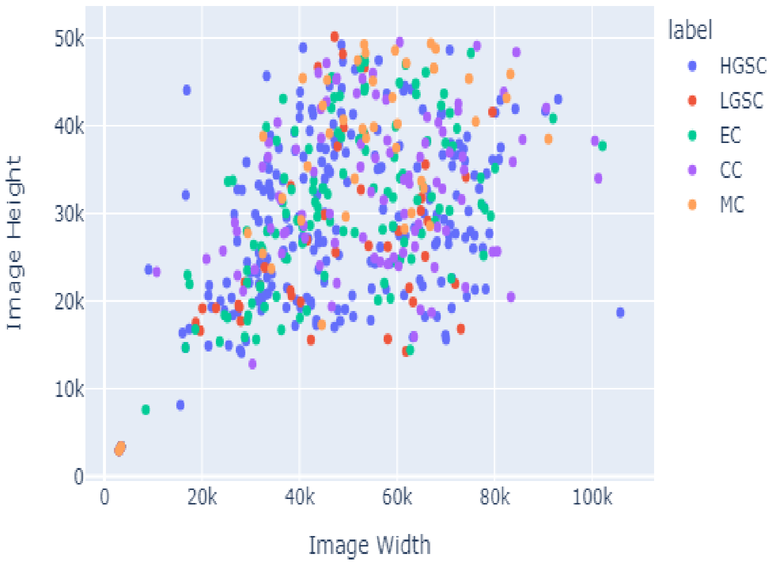


Fig. 4: image width and height of all sub types of cancer

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

Fig. 3 shows the image width and height of sub-types of cancer in different colours, and it also shows the label of cancer.

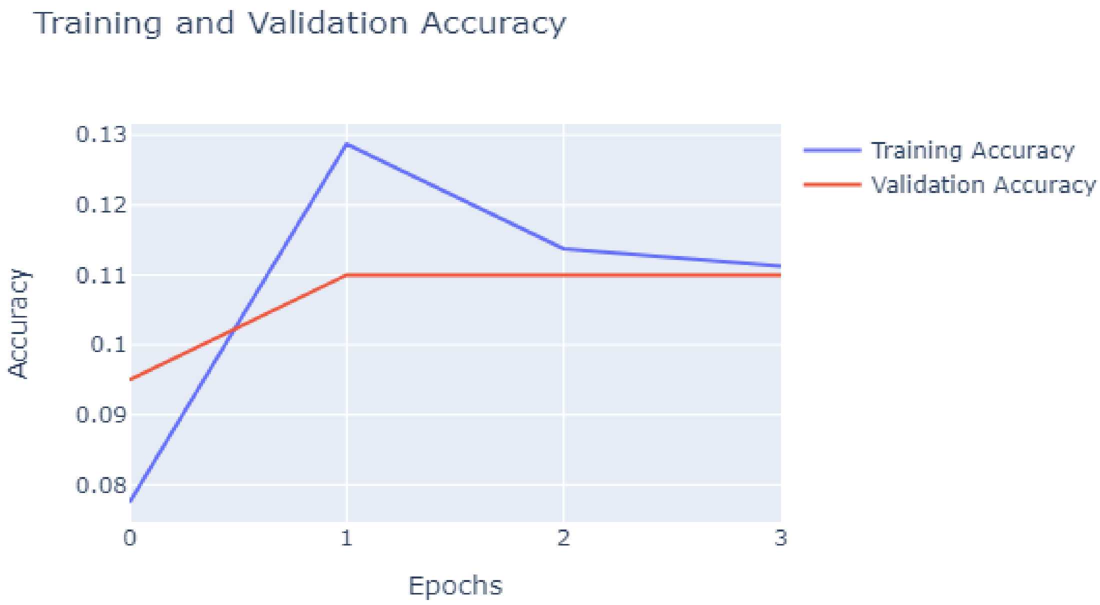


Fig .5: Training and Validation Accuracy

The Fig. 4 plot displays the training and validation accuracy over epochs for a neural network trained with the best model. The model achieves satisfactory accuracy, demonstrating effective learning and validation. The training accuracy value has a high value compared to validation accuracy.

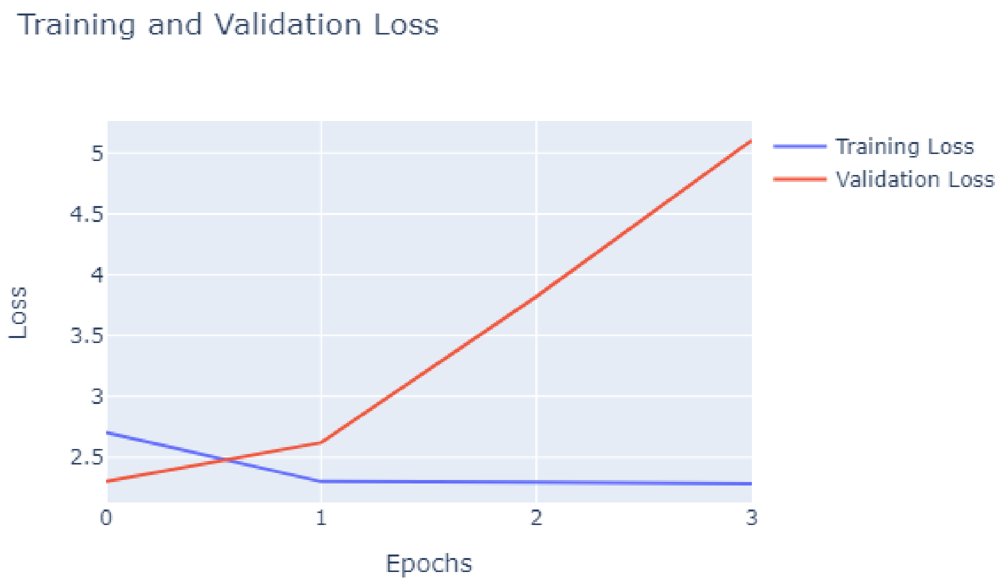


Fig .6: Training and Validation Loss

The Fig. 5 The plot illustrates the training and validation loss over epochs for a neural network using the object. The model effectively minimizes loss, indicating successful training and validation. The validation loss is greater when compared to the training loss of data.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(06), 3212-3220 | ISSN: 2251-6719

Out[35]:

	image_id	label
0	41	HGSC

Fig .7: Detection of sub class type cancer

In [37]: 1 submission

Out[37]:

	image_id	label
0	41	3

Fig .8: Submission file and label id

Figs. 6 and 7 show the prediction of a subtype of cancer label and label number.

5. CONCLUSION

In conclusion, the UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN) harnesses advanced convolutional neural networks (CNNs) to adeptly classify diverse ovarian cancer subtypes. This innovative approach overcomes challenges in traditional methods by leveraging CNN's hierarchical feature learning. UBC-OCEAN integrates outlier detection, enhancing its capability to identify rare subtypes. The proposed CNN model, with its strategic architecture, signifies a paradigm shift in ovarian cancer research, offering a promising avenue for precise classification. The fusion of deep learning techniques, especially CNNs, underscores a transformative step towards improved diagnostic accuracy and patient outcomes in ovarian cancer.

REFERENCES

1. Hema, L. K., R. Manikandan, Majid Alhomrani, N. Pradeep, Abdulhakeem S. Alamri, Shakti Sharma, & Musah Alhassan. (2020). Region-Based Segmentation andClassification for Ovarian Cancer Detection Using Convolution Neural Network. *Contrast Media & Molecular Imaging*, 12.
2. Ponnada, V. T., & S.V. Naga Srinivasu. (2020). Efficient CNN for Lung Cancer Detection. *International Journal of Recent Technology and Engineering (IJRTE)*, 5.
3. ZHANG, Z., & YIBO HAN. (2020). Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier With an Advanced Machine Learning Approach. *SPECIAL SECTION ON DEEP LEARNING ALGORITHMS FOR INTERNET OF MEDICAL THINGS IEEE ACCESS*, (p. 10).
4. Kaur, B., Kulvinder Singh Mann, & Manpreet Kaur Grewal. (2017). Ovarian Cancer stage based detection on Convolutional Neural network. *Proceedings of the 2nd International Conference on Communication and Electronics Systems*, (p. 5).
5. Mathur, M., Vikas Jindal, & Gitanjali Wadhwa. (2020). Detecting Malignancy of Ovarian Tumour using Convolutional Neural Network: A Review. *2020 Sixth International Conference on Parallel, Distributed and Grid Computing*, (p. 6).
6. Yang Yu, Q. Y., & Liu Shaohan. (2020). An Improved Faster R-CNN for Colorectal Cancer Cell Detection. *2020 IEEE the 3rd International Conference on Electronics and Communication Engineering*, (p. 5).
7. Nagarajan, P. H., & N. Tajunisha. (2023). OPTIMAL PARAMETER SELECTION-BASED DEEP SEMI-SUPERVISED GENERATIVE LEARNING AND CNN FOR OVARIAN CANCER CLASSIFICATION. *ICTACT JOURNAL ON SOFT COMPUTING*, 7.
8. Reda, I., Babajide O., Ayinde, , Mohammed Elmogy, Ahmed Shalaby, Moumen El-Melegy, . . . Ayman El-Baz. (2018). A NEW CNN-BASED SYSTEM FOR EARLY DIAGNOSIS OF PROSTATE CANCER. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (p. 4).
9. WANG, Z., MO LI, HUAXIA WANG, HANYU JIANG, & YUDONG YAO. (2019). Breast Cancer Detection Using Extreme Learning Machine Based on Feature Learning Machine Based on Feature., (p. 13).