*DOI: 10.15849/IJASCA.230720.04*

*Int. J. Advance Soft Compu. Appl, Vol. 15, No. 2, July 2023*
*Print ISSN: 2710-1274, Online ISSN: 2074-8523*


# Colorectal Cancer Prediction using ResNet-CNN Classification Method

**V. Gokula Krishnan, P. A. Abdul Saleem, K. Sathyamoorthy, K. Hema Priya and TK. Senthil Kumar**

Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai, Tamil Nadu, India
e-mail: gokul_kris143@yahoo.com
Professor, Department of CSE (DS), CVR College of Engineering, Mangalpally, Hyderabad, Telangana, India
e-mail: drsaleemprincipal@gmail.com
Assistant Professor, Department of CSE, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu, India
e-mail: pitsathyamoorthy@gmail.com
Assistant Professor, Department of CSE, Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India
e-mail: hemu.june3@gmail.com
Senior Data Scientist, NYBL, Dubai, UAE
e-mail: tkseneee@gmail.com

**Abstract**

***The common cancers with extraordinary death rates are colorectal cancer (CRC). The determination of microsatellite instability (MSI) is imperative for the analysis of the associated disease and the determination of treatment. We were driven to investigate a model for the classification of CRC images by the outstanding presentation of the Convolutional Neural Networks (CNN). In this study we propose an enhanced CRC image clustering residual network. The pipeline proposed comprises the serial construction between the segmentation system and the classification system. The first phase segments the tumor area with the Feature Pyramid network (FPN) from the pathological image. Total widespread variation Fuzzy C-means clustering (TGVFCMS) is employed in this proposed system for segmentation purposes. The dataset used in the study included the PAIP 2020 test to the CRC MSI-H rating. The proposed result shows that the proposed ResNet provides much better outcomes compared to other Deep learning techniques.***

**Keywords**: *Colorectal cancer, Deep learning, Feature Pyramid Network (FPN), PAIP 2020 challenge and Convolutional Neural Network (CNN).*

## 1 Introduction

A malignant tumor in the rectum or colon that is present in several countries is CRC, a rapidly growing cancer related to death. According to National Cancer Centre Korea, in 2017, 28,111 individuals were diagnosed with CRC; in addition, that year, 8,691 people died alone in Korea [1] due to CRC. In 2020, over 140 000 patients are recently diagnosed

***Received 23 March 2022; Accepted 29 May 2023***

with CRC and over 50, 000 are nationally diagnosed as a result of the condition [2]. In the past few years, however, the incidences and mortality rates showed degrading trends due to the agreement on the molecular classification and its targets [3]. CRC was usually the leading five cancers in the field of mortality rates.

Typical pathogenesis of CRC, chromosoma instability (CIN) and methylator phenotype of CpG Island is microsatellite instability (MSI). The increase in anomalous genetic changes resulting from the damaged repair system of DNA mismatch which is detectable molecularly on repeated motifs of DNA [4]. The Society of National Cancer defines CRC tissue MSI status as MSI-H. If five irregular microsatellite markers are present, MSI-L if one, or MSI-Stable if none. The MSI is essential to identify the status of the disease, for example Lynch syndrome, and to improve cancer treatment prognosis. Therefore, the position of MSI in the tissue is strongly recommended to pathologists. The haematoxylin and eosine-based CRC tissue is examined and any irregularities in the microscopic images are manually checks. This technique, however, is fatiguing and occasionally mistaken [5].

The tumor microenvironment has been deeply attracted by the closely related development and progression of tumors. In the development of colon cancer, interactions among epithelial, stromal and immune cells are key features of prognostic and therapeutic strategies [6-7]. Histopathological images taken from tissue slides comprise a wealth of information on cell morphology and the structure of tissues which is a cost-effective and readily-accessible tool in the clinical decision making of tumor environments. For example, H&E-driven histology slides can help pathologists identify cell categories such as tumor cells and stroma for their diagnosis. Manual inspection, however, requires less time and effort and may not capture patterns. In current years, the advanced clinical diagnosis was focused on artificial intelligence (AI), which significantly expands traditional pathology capacity. Deep learning, which has achieved excellent human diagnosis performance in numerous specific aspects including image-based cell discovery and classification [8-9], is a representative example. Digital pathology based on AI cannot only accurately distinguish between various tissue and cell type [10-11]; image-based features can also be linked with molecular features such as gene mutations, MIs and molecular subtypes [12–14].

# 2 Related Work

In this section, a study of preceding systems that are used to detect the CRC is presented. In addition, the achievement of the existing techniques along with its drawback also given in the below paragraph.

Hu et al., [15] aimed at identifying potential, in circulatoric exosomes, long, non-coding RNAs (lncRNAs) for stage CRC detection. Ultrasonophagation was surveyed by extraction of the total RNA by TRIzol reagent from plasma of patients with CRCs (n = 50) and healthy characters (n = 50, respectively). The exosomal lncRNA profleting analysis was used in both groups and a retro-transcription, quantitative reverse polypherase chain (qRT-PCR) reaction was used in real-time patient's subjects to determine expression levels of lncRNAs. Although positive results have been obtained, further various studies are needed to clarify the particular role of these lncRNA in the CRC.

To assess and authenticate individual plasma proteins, Bhardwaj, et al., [16] was to be signed for early detection of CRC. Proteins were measured during three-stage designs by

liquid chromatography with (LC - MRM-MS) and then by a nearness extension test in a finding set consisting of 96 CRC cases recently diagnosed and 94 neoplasm-free neoplasms in colonoscopy screening. The Lasso regression derived two algorithms (one for each measuring method). In addition, another algorithm, the maximum talented protein marker in PEA measurement that wasn't available from the LC/MRM-MS measures, plus ampheglin has been made of the same eleven biomarkers. In this study the examples were collected, processed and stored in the three studies, even though they were selected from 3 different trials, with the same standardized operating procedure (SOPs).

Herreros-Villanueva, [17] in the development of a robust predictive technique to distinguish healthy persons of CRC or advanced adenomas (AA) illnesses, examined the presentation of particular samples. 297 patients from 8 Spanish centers, with 100 healthy persons, 101 with an AA diagnosis and 96 CRC cases were the undertaken case. In order to quantify a miRNA signature in plasma samples, quantitative reverse transcript was used in real time. For the best predictive model, binary classifiers (linear SVM support [SVM], radial SVM and polynomial SVM) were developed. A signature of 6 miRNAs was identified and validated as predictors that can expressively distinguish between CRC and AA patients and healthy patients. However, major validation studies should be conducted in asymptomatic test participants.

Zonta, et al.,[18] The Protocol was studied and proposed that a device made up of chemoresistive gas sensor systems made of semiconducting materials be validated clinically, so that the alteration among fecal sigh of healthy individuals and subjects of high-risk colorectal polyps or tumors can be identified. Tests were associated to the fecal occult blood test results and gold standard colonoscopy.

The four studies included in Duran-Sanchon, et al., [19]. During the discovery phase, 124 fresh, and non-tumor examples (30 CRC; 32 AAs) have been profiled throughout Spain by miRNA expression (30 CRC). In the technical stage of validation miRNAs with altered tumor versus non-tumor tissue expressions were quantified in the fecal samples of a subgroup of discovery patients (n=39) and persons with no colorectal neoplasms (controls, n=39) with reverse transcription polymerase chain response. In the miRNAs of patients. Finally, a model was developed to classify patients with advanced neoplasms (CRC or AAs) using colonoscopic findings as the reference standard based on their miRNA profiles. However, it is not possible to compare these results directly with those obtained by participating in test-native programs in a limited study, which is the use of FIT positive persons to evaluate the accuracy of fecal miRNA signature.

Future efforts to reduce CRC morbidity and death will depend on the techniques for CRC management and the implementation of worldwide screening programmes, both of which were introduced by Xi Y et al., [20]. Prognosis and therapy response in CRC vary depending on the patient's subtype. The accuracy of a method to categorize CRC is crucial for both fundamental research and clinical result. To better understand the molecular basis

of CRC, we cover the various CRC subtypes, and we sum up the current risk factors, preventive, and screening measures for CRC.

According to Sirinukunwattana K et al. [21] Create a Using data from three separate datasets (training on FOCUS trial, n=278 patients; test on rectal cancer biopsies, GRAMPIAN cohort, n=144 patients; and The Cancer Genome Atlas (TCGA), n=430 patients), a neural network was trained and evaluated. By comparing the predictions from the CMS classifier using a random forest against those using a single sample, the true CMS calls were determined. corresponding to the bioinformatics analysis of molecular information. imCMS reproduced predicted relationships with genomic and epigenetic changes and exhibited comparable prognostic linkages as transcriptomic CMS, while also correctly classifying samples that could not be categorized by RNA expression profiling.

# 3 Proposed System

A model for CRC image classification has been motivated by the outstanding performance of CNN in image recognition. In this study we propose an enhanced CRC image clustering residual network. Figure 1 illustrates the proposed framework and explains them in detail in the subsections that follow.

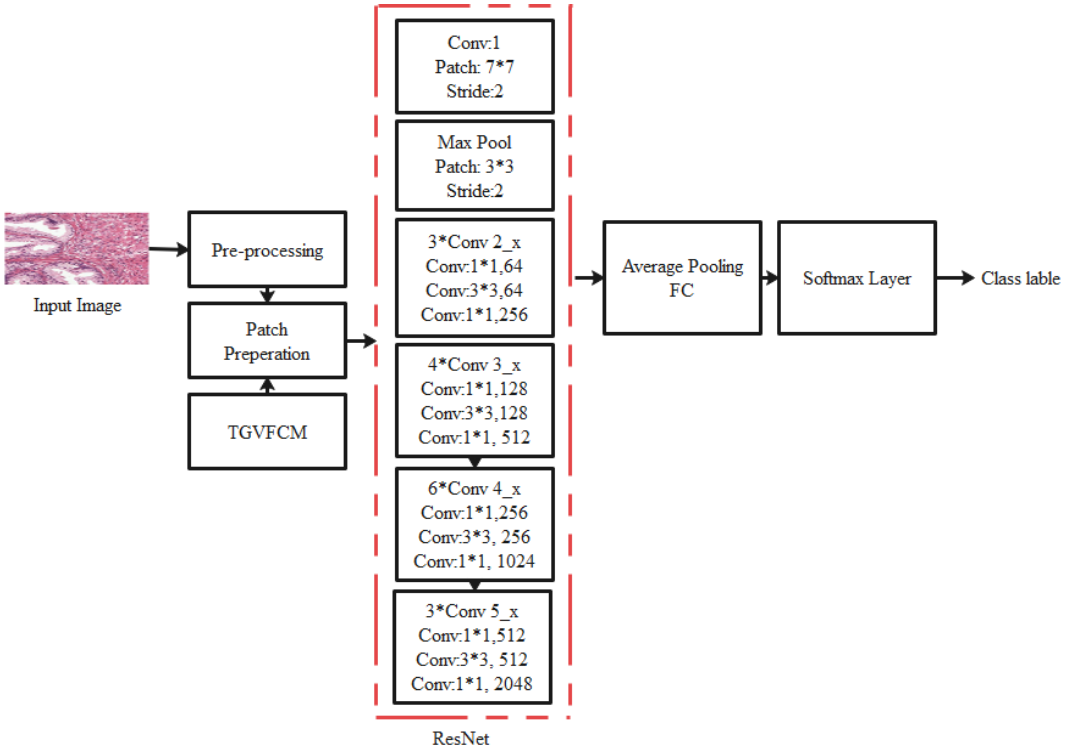


Figure 1. Projected Flow Diagram

## 3.1. Dataset

The dataset as PAIP 2020 is used in the experiment to analysis the CRC MSI-H rating. The dataset consist of totally 10 validation sets with 47 training sets. The datasets for validation are accepted from the course and used only for authentication. Each dataset has WSIs, XML files that annotate a tumor zone and MSI-classification ground truths. The WSIs in the dataset are resected tissue microscopic images with hematoxylin and eosin coloring and are scanned by APIRIO AT2 at 40 times magnification. The WSIs also comprised low

magnification images as 20 and 10 digits, to show the context information on the original 40-litre-larging image. In this study, WSIs are mostly used in terms of time and efficiency with 20 kilograms and 10 kilograms magnifications. At least one colorectal tumor region is included in all WSIs. Figure 2 shows some of the sample data-sets.

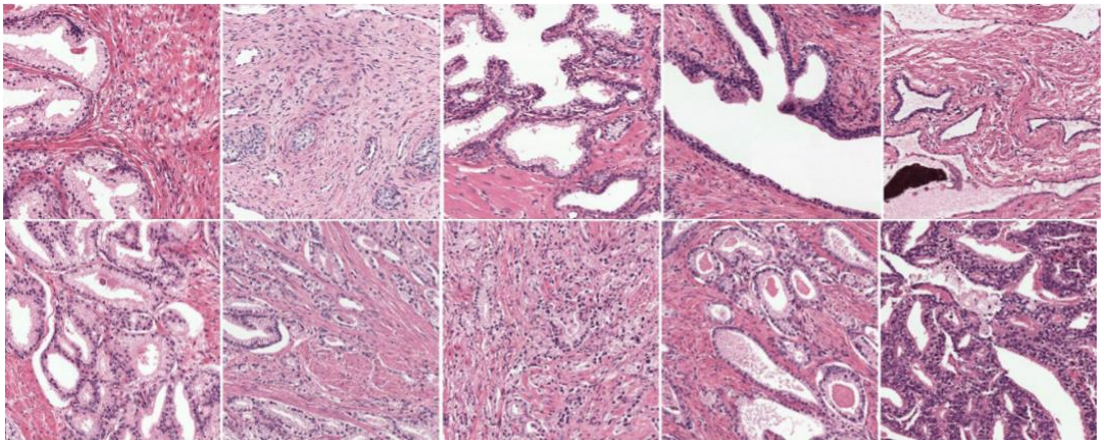

Figure 2. Sample PAIP 2020 images

## 3.2. Pre-Processing

The quality and number of data sets available depend largely on the presentation of CNN models. The processing of CRC images we have already played an important role in image pre-processing and enhancement methods. Different noise types in oral images are removed by pre-processing. Improvement of the image is used to expand datasets, increase accuracy and reduce the overfitting of the models. In this method we employed image enhancement techniques using a single sample, to create multiple CRC image versions through color, transformation (translation, Scaling and Rotation) and rotation to prevent overfitting of the model. Then, stain normalization is used to normalize CRC images and remove color and intensity variations of data sets, which improve the predictive accuracy before network training.

### 3.2.1. Patch Preparation

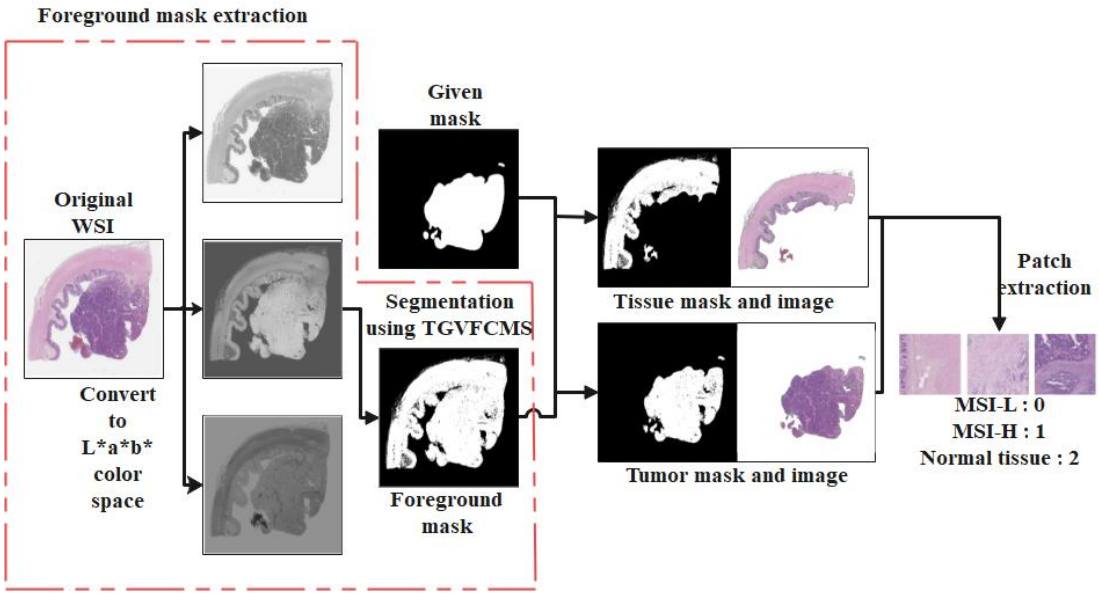

Figure 3. Patch preparation procedure overview: new tissue and mask formation, foreground mask extraction, and patch extraction

We have adapted the patch extraction method, taking into consideration the color properties of the discolored tissue. Figure 3 describes the overall process of patch preparation. The operation started with the extraction of the front mask. The original WSI color format is transformed from RGB to the CIE $L * a * b *$ color space to help separate the tissues in red from the circumstantial area. To remove the foreground mask, we have used a* image channel and a locked morphology to reduce noise by the TGVFCMs segmentation method. This foreground mask is used to make masks for the normal and tumor area using the XML annotation file. Apply pixel NOT and AND operations on the two masks are used to generate the tissue and tumor masks. The mask gotten from the marked XML file, as shown in Figure 4, included the background. This could impair network performance and, as shown in Figure 4.d, we used an afresh created mask to develop precision.

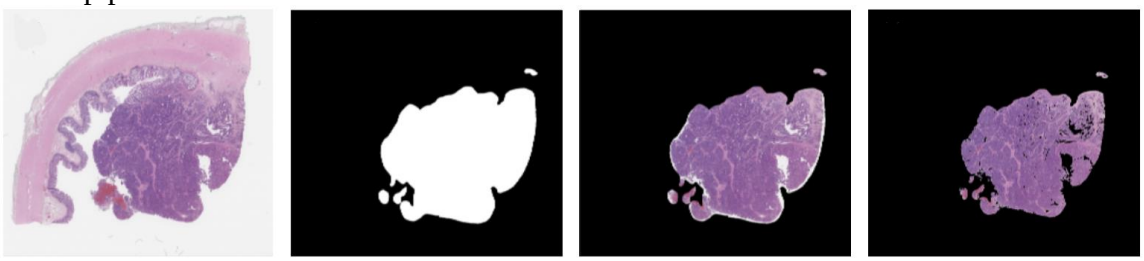

Figure 4. The unique (WSI) and the related tumor mask: (**a**) the unique WSI, (**b**) the tumor mask produced from the specified XML footnote, (**c**) the segmented tumor part from WSI, and (**d**) the formed tumor mask without contextual.

## 3.3. Segmentation using TGVFCMS

The capacity of the TGVFCMS technique to withstand noise while still preserving edges is crucial. In particular, the TGV's sought-after anti-aliasing capabilities make it a valuable instrument for gauging image possessions including noise suppression and the maintenance of crisp edges. In this study, we modified the TGV throughout the smoothing phase of our TGVFCMS to eradicate the background noise and artefacts that plague FCM-based methods. Here is a quick breakdown of how the TGV is built:

$$TGV_a^k(u) = sup\left\{\int_\Omega udiv^k v dx \mid v \in \mathbb{C}_c^k\left(\Omega, Sym^k(\mathbb{R}^d)\right), ||div^l v||_\infty \leq a_l\right\} \quad (1)$$

Where $l = 0,1, \ldots, k-1,$ and $k \in \mathbb{N}$ specifies a TGV, and $a = (a_0, a_1, \ldots, a_{k-1})$ indicates the positive weight to TGV. $Sym^k(\mathbb{R}^d)$ Indicates the space of symmetric $k$-tensors. For every component $\eta \in M_{k-1}$,the l-divergence of the symmetric k tensor is assumed by

$$(div^l v)_\eta = \sum_{r \epsilon M_1} \frac{l!}{\gamma!} \frac{\partial^l v_{\eta+\gamma}}{\partial x^\gamma} \quad (2)$$

Where $M_k$is the multi-index of order $k$

$$M_k = \left\{\eta \epsilon \mathbb{N}^d \mid \sum_{i=1}^d \eta_i = k\right\} \quad (3)$$

The ∞- norm for the symmetric k-vector arena is given as

$$||v||_\infty = \underset{x \epsilon \Omega}{sup}\left\{\left(\sum_{\eta \epsilon M_k} \frac{k!}{\eta!} v_\eta(x)^2\right)\right\} \quad (4)$$

The two gradients are gotten in (1) only essential to be rare, which can strikingly lessen the scale artefacts.

Here, we take TGV the second-order consider as.

$$TGV_a^2(u) = sup\left\{\int_\Omega udiv^2 vdx \,|v \in \mathbb{C}_c^2(\Omega, S^{d\times d}), \left||v|\right|_\infty \le a_0, ||div\ v||_\infty \le a_1\right\} \quad (5)$$

Where $\mathbb{C}_c^2(\Omega, S^{d\times d})$ Expresses an efficiently space below the set of symmetric metrics $S^{d\times d}$. In specific, the consistent meanings of difference and criteria can be intended as shadows:

$$(div\ v)_i = \sum_{j=1}^d \frac{\partial v_{ij}}{\partial x_j}, (div^2\ v)_i = \sum_{j=1}^d \frac{\partial^2 v_{ii}}{\partial x_i^2} + 2\sum_{i<j} \frac{\partial v_{ij}}{\partial x_j \partial x_j} \quad (6)$$

and

$$||v||_\infty = \sup_{x\ \in\ \Omega}(\sum_{i=1}^d |v_{ii}(x)|^2 + 2\sum_{i<j} |v_{ij}(x)|^2)^{1/2}$$

$$||div\ v||_\infty = \sup_{x\ \in\ \Omega}\left\{\sum_{i=1}^d \left|\sum_j^d \frac{\partial^2 v_{ij}}{\partial x_j^2}(x)\right|^2\right\}^{1/2} \quad (7)$$

If the negligible solution for arenas is taken on Ω and $\varepsilon(v) = (\nabla v + \nabla u\, T)/2$ gives the symmetrized derived. Here definition (7) demonstrates that the smoothed regions receive less than u = v from u = u2. Minimization for v = 0 occurs closest to the boundary, where 2 u is "larger" than u. Hence, it allows for a stable positive weight-to-negative-weight ratio between the derivatives. The two weights, _0 and _1, are initialized to 0.1 and 0.15, respectively. As a consequence of calculating the second-order TGV, the predicted TGVFCMS can produce results that are more robust in the face of noise and factor retention [20].

## 3.4. CNN Architecture

In practice, different activation and loss function, optimization of the parameters, regularization, and structural innovations have been explored by CNN architectures. In this study, we have tried to adapt our specific data and requirements to find the right configuration. We regarded the classical ResNet architecture that has achieved excellent image classification performance through the direct learning of useful structures from the image patches and the loss function an optimization. The main disadvantage of the CNN architecture is that due to max pooling operations, no account is taken of the location and spatial relations between the images. But the relationship between small objects plays an important role in histopathology images.

This construction introduced the concept called Residual Network to solve the problem of the disappearance/exploding gradient. The modular unit of the widespread residual network construction is a generalized residual block containing of parallel states for a residual current, $r$, containing identity shortcuts and resembles the residual block structure of the residual one-columned ResNet (parameters $W_l, r \to r$), and a transient stream, t, which is a standard convolutional layer ($W_l, t \to t$). Two additional sets of conv filters in each block ($W_l, r \to t, W_l, t \to r,$) also transfer information across streams.

$$r_{l+1} = \sigma(conv(r_l, W_{l,r\to r}) + conv(t_l, W_{l,t\to r}) + shortcut(r_l)) \quad (8)$$

$$t_{l+1} = \sigma(conv(r_l, W_{l,r\to t}) + conv(t_l, W_{l,t\to r}))$$

Before using batch Normalization and ReLU non-linearities (together μ), the same-stream and cross-stream activations are added to the shortcut connection for the residual flux (Equation 8). Rest stream function $r$ is similar to the original ResNet structure with shortcut connects between each processing unit, while transient stream t enables nonlinear, non-linear processing, of information from any stream, without shortcut connection, to be discarded. The shape of a shortcut can be an identity function with a suitable padding or projection. As one convergence layer, we tool the generalized residual block with an improved initialization called ResNet Unit.

Either a normal CNN layer (with a residual stream being zeroed) or a single-layer ResNet block can serve the generalized residual block. The widespread residual architecture is able to learn between these things, including the standard 2-layer ResNet block, by repeating several times the widespread residual block. The architecture allows the network to learn residuals with a variable efficient number of processing steps before adding back to the residual stream. The generalized residual block is not CNN specific and can be used for fully connected standard layers and other feedforward layers. Substitute each conv layer in a residual block of the original ResNet with a widespread residual block leads us to a new ResNet architecture.

# 4 Results and Discussion

Matlab with an Intel i3 processor at 3.0 GHz, a 1 TB hard drive, and 8 GB of Memory is used to evaluate the proposed system. The PAIP 2020 challenge dataset is detailed below, and its performance is evaluated by comparing the suggested scheme with the conventional system. We created patches for three time periods by pre-processing the training dataset (e.g., MSI-L, MSI-H and normal tissue). To avoid the class imbalance problem, classes with fewer patches were oversampled. The remaining 20% are used for cross-validation and the rest for additional training. Patches undergo transformations including vertical and horizontal flips in addition to other types of data increases to prevent over position.

## 4.1. Performance Metrics

The following equations represent various parametric measures used to validate the proposed system's performance: accuracy, sensitivity, and specificity.

$$SE = \frac{tp}{tp + fn} \tag{9}$$

$$SP = \frac{tn}{tn + fp} \tag{10}$$

$$AC = \frac{tp + tn}{tp + fp + tn + fn} \tag{11}$$

$$FM = \frac{tp}{tp+1/2(fp+fn)} \tag{12}.$$

## 4.2. Evaluation

In this study, we used a newly created online dataset for validating the performance of the proposed algorithm. But the existing techniques used collected samples from different

hospitals, hence these procedures are applied on our online dataset for authentication procedure.

Table 1. Comparative analysis of Projected CNN with numerous Deep Learning Procedures

| | **Accuracy** | **Sensitivity** | **Specificity** | **F-Measure** |
|---|---|---|---|---|
| CNN- ResNet | 95.00 | 100 | 90.00 | 95.24 |
| **LSTM** | 77.50 | 75.00 | 80.00 | 76.92 |
| **Auto encoder** | 80.00 | 85.00 | 75.00 | 80.95 |
| **Recurrent** | 75.00 | 80.00 | 70.00 | 76.19 |
| **Recursive** | 85.00 | 90.00 | 80.00 | 85.71 |
| **GAN** | 87.50 | 90.00 | 85.00 | 87.80 |

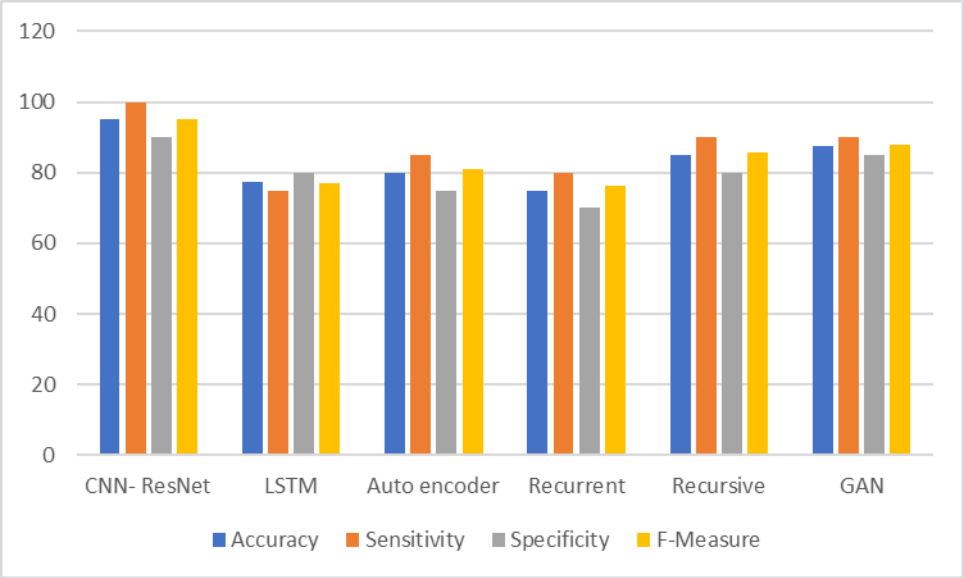


Figure 5. Graphical representation of Projected CNN with numerous Deep Learning Procedures

The presentation of the projected CNN is associated with that of other deep learning approaches using a variety of parametric parameters, as shown in Table 1 and figure 5. The F-measure for LSTM was 76.92% accurate, exceeding the accuracy goal of 77.50%. Then, the Auto encoder obtained an F-measure value of 80.95%, indicating a classification accuracy of 80.00%. With an f-measure of 76.19%, the recurrent model performed at an accuracy of 75.00%. Both the recursive model (sensitivity = 90.00%) and the GAN model (accuracy = 87.50%, specificity = 85.00%) were successful. The proposed CNN, however, outperformed the baseline by a significant margin (95.00% against 95.24% F-Measure). According on the results of this evaluation, the suggested CNN-ResNet outperformed the other deep learning models.

Table 2. Relative investigation of Proposed CNN with numerous CNN Constructions

| **Metrics** | **ResNet** | **AlexNet** | **LeNet** |
|---|---|---|---|
| Accuracy | 95.00 | 85.00 | 80.00 |
| **Sensitivity** | 100 | 80.00 | 70.00 |
| **Specificity** | 90.00 | 90.00 | 90.00 |
| **F-Measure** | 95.24 | 84.21 | 77.78 |

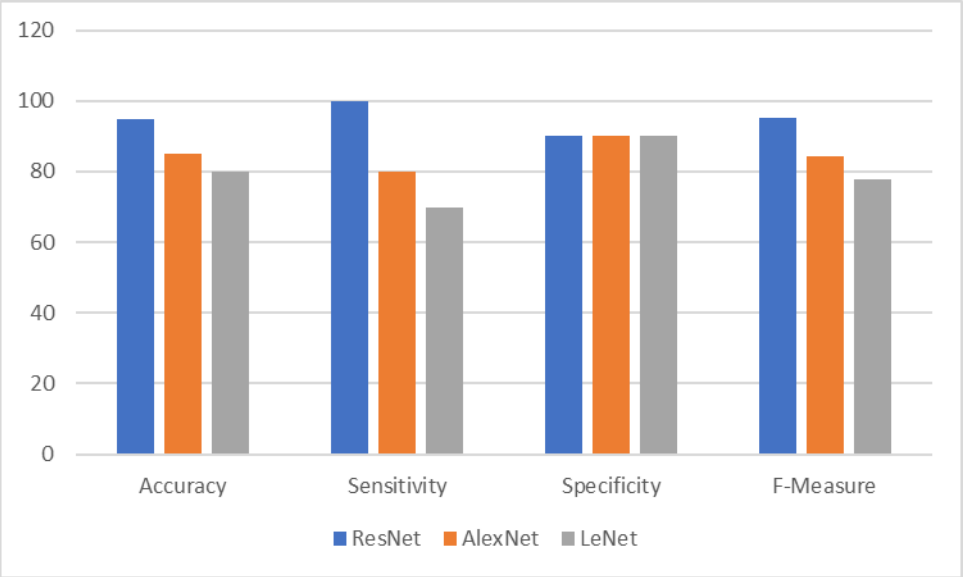


Figure 6. Graphical representation of Proposed CNN with numerous CNN Constructions

Table 2 and figure 6 displays the results of a comparison between the proposed CNN and other CNN architectures. AlexNet, LeNet, and ResNet are the three paradigms of network architecture. Using AlexNet, we were able to get an F-measure value of 84.21% and an accuracy of 85.0%. The F-measure for the LeNet model was 77.78%, and its accuracy was 80.00%. In the end, ResNet's model performed better statistically, scoring an F-measure value of 95.24% and an accuracy of 95.00%. Conclusion: The ResNet model outperformed the competition in terms of presentation quality.

# 5 Conclusion

A typical multi-class classification method simultaneously divides three MSI statuses; however, the difference between MSI-H and other tissue kinds is poorly defined. In this paper, the MSI status of CRC tissues is predicted with a novel double-stage classification pipeline. Two classifiers are serially connected: the classification process by the proposed classification pipeline. In the first phase, the segmentation network prevented the improper prediction of MSI-L normal tissue, where TGVFCMS is employed in this proposed system for segmentation purposes and multi-classification technique is used for final prediction. The dataset used in the study included the PAIP 2020 challenge to the CRC classification of MSI-H rating. The proposed result shows that the proposed ResNet provides better classification results compared to other Deep learning techniques.

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

**Data Availability Statement**
The data shall be made available on request.

**Funding Statement**
This research work is self-funded.

# References

[1]. Hong S, Won YJ, Park YR, Jung KW, Kong HJ, Lee ES (2020) Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2017. Cancer research and treatment: official journal of Korean Cancer Association. 52(2):335.

[2]. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. CA: a cancer journal for clinicians. 66(1):7-30.

[3]. Jass JR (2007) Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology. 50(1):113-30.

[4]. Bardhan K, Liu K (2013) Epigenetics and colorectal cancer pathogenesis. Cancers. 5(2):676-713.

[5]. Konukiewitz B, Schmitt M, Silva M, Pohl J, Lang C, Steiger K, Halfter K, Engel J, Schlitter AM, Boxberg M, Pfarr N. Loss of CDX2 in colorectal cancer is associated with histopathologic subtypes and microsatellite instability but is prognostically inferior to hematoxylin–eosin-based morphologic parameters from the WHO classification. British journal of cancer. 2021 Dec 7;125(12):1632-46..

[6]. Naschberger E, Liebl A, Schellerer VS, Schütz M, Britzen-Laurent N, Kölbel P, Schaal U, Haep L, Regensburger D, Wittmann T, Klein-Hitpass L (2016) Matricellular protein SPARCL1 regulates tumor microenvironment–dependent endothelial cell heterogeneity in colorectal carcinoma. The Journal of clinical investigation. 126(11):4187-204.

[7]. Jackstadt R, van Hooff SR, Leach JD, Cortes-Lavaud X, Lohuis JO, Ridgway RA, Wouters VM, Roper J, Kendall TJ, Roxburgh CS, Horgan PG (2019 Sep 16) Epithelial NOTCH signaling rewires the tumor microenvironment of colorectal cancer to drive poor-prognosis subtypes and metastasis. Cancer cell. 36(3):319-36.

[8]. Javed S, Mahmood A, Fraz MM, Koohbanani NA, Benes K, Tsang YW, Hewitt K, Epstein D, Snead D, Rajpoot N (2020 Jul 1) Cellular community detection for tissue phenotyping in colorectal cancer histology images. Medical image analysis. 63:101696.

[9]. Sirinukunwattana K, Raza SE, Tsang YW, Snead DR, Cree IA, Rajpoot NM (2016 Feb 4) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE transactions on medical imaging. 35(5):1196-206.

[10]. Cheng J, Mo X, Wang X, Parwani A, Feng Q, Huang K (2018 Mar 15) Identification of topological features in renal tumor microenvironment associated with patient survival. Bioinformatics. 34(6):1024-30.

[11]. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, Gaiser T, Marx A, Valous NA, Ferber D, Jansen L (2019 Jan 24) Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS medicine. 16(1):e1002730.

[12]. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira AL, Razavian N, Tsirigos A (2018) Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine. 24(10):1559-67.

[13]. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, Marx A, Boor P, Tacke F, Neumann UP, Grabsch HI (2019) Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature medicine. 25(7):1054-6.

[14]. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, Leedham SJ, Chatzipli A, Hardy C, Whalley CM, Wu CH (2021 Mar 1) Image-based

consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 70(3):544-54.
[15]. Hu D, Zhan Y, Zhu K, Bai M, Han J, Si Y, Zhang H, Kong D (2018) Plasma exosomal long non-coding RNAs serve as biomarkers for early detection of colorectal cancer. Cellular Physiology and Biochemistry. 51(6):2704-15.
[16]. Bhardwaj M, Gies A, Weigl K, Tikk K, Benner A, Schrotz-King P, Borchers CH, Brenner H (2019) Evaluation and validation of plasma proteins using two different protein detection methods for early detection of colorectal cancer. Cancers. 11(10):1426.
[17]. Herreros-Villanueva M, Duran-Sanchon S, Martín AC, Pérez-Palacios R, Vila-Navarro E, Marcuello M, Diaz-Centeno M, Cubiella J, Diez MS, Bujanda L, Lanas A (2019) Plasma MicroRNA signature validation for early detection of colorectal cancer. Clinical and translational gastroenterology. 10(1).
[18]. Zonta G, Anania G, Feo C, Gaiardo A, Gherardi S, Giberti A, Guidi V, Landini N, Palmonari C, Ricci L, de Togni A (2018 Jun 1) Use of gas sensors and FOBT for the early detection of colorectal cancer. Sensors and Actuators B: Chemical. 262:884-91.
[19]. Duran-Sanchon S, Moreno L, Augé JM, Serra-Burriel M, Cuatrecasas M, Moreira L, Martín A, Serradesanferm A, Pozo À, Costa R, Lacy A (2020 Mar 1) Identification and validation of microRNA profiles in fecal samples for detection of colorectal cancer. Gastroenterology. 158(4):947-57.
[20]. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Translational oncology. 2021 Oct 1;14(10):101174.
[21]. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, Leedham SJ, Chatzipli A, Hardy C, Whalley CM, Wu CH. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 2021 Mar 1;70(3):544-54.
[22]. Zhang X, Pan W, Wu Z, Chen J, Mao Y, Wu R (2020 May 19) Robust Image Segmentation Using Fuzzy C-Means Clustering with Spatial Information Based on Total Generalized Variation. IEEE Access. 8:95681-97.

**Notes on contributors**

**Dr. V. Gokula Krishnan** is currently working as Professor in the Department of Computer Science and Engineering in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India. He has completed his Under-Graduation (BE) in Anna University, Post-Graduation (M.Tech) in Dr. MGR University and Ph.D in Sathyabama Institute of Science and Technology, Chennai. He has more than 16 years of teaching experience. He has published several papers in SCI/Scopus/WoS indexed journals and also he has presented various papers in National/International Conferences. His area of interest includes Computer Networks, Computer Architecture, Data Structures, and Software Engineering etc. He serves as the guest editor, editorial member and also as reviewer in many reputed international journals. He is a member of Professional Bodies like ISTE, IAENG, CSI and IEEE etc.

**Dr. P. A. Abdul Saleem** is presently working as Professor – Computer Science and Engineering (DS) , CVR College of Engineering, Ibrahimpatnam, Ranga Reddy District. Telangana State, India. He has completed UG as well as PG from University of Madras and Doctorate from University of Allahabad. He has 29 Years of Industrial and Teaching with Research Experience in the fields of Electronics, Communication and Computer Technologies. His Research Specialization is Wireless Mobile ADHOC Networks and also interested in the area like MANET Security, Machine Learning, Computer Architecture, Bio-Informatics, Artificial Intelligence etc.

K. Sathyamoorthy is currently working as Assistant Professor in the Department of Computer Science and Engineering in Panimalar Engineering College, Chennai, Tamil Nadu. He has obtained his Under-Graduation (B.Tech) from Anna University, PostGraduation (M.Tech) from Sathyabama University and he is perusing his PhD in Vel Tech University, Chennai. He has more than 12 years of experience in the industry and teaching field, and he has published many papers in national/international conferences and journals. Her area of interest includes Data Structures, Computer Networks and Image Processing etc.

Mrs. K. Hema Priya works as Assistant Professor, Department of Computer Science and Engineering in Easwari Engineering College. She is pursuing her Ph.D. in Anna University, completed her PG (M.Tech) in SRM University and UG (BE) in Anna University. She has around 10 years of experience in teaching. Her area of interest includes Big Data, Machine Learning, Deep Learning, Data Mining. She has published several papers in Scopus/SCI/WOS indexed journals and also presented papers in various International Conferences.

Dr Senthilkumar TK obtained his Bachelor's degree in Electrical and Electronics Engineering and his Master's degree in Applied Electronics from Anna University, Chennai, India. He completed his research thesis on the area of Medical Image Processing in the Department of ICE, Anna University. At present, he is working as a Senior Data Scientist in the department of Computer Vision at NYBL. His specialization includes Machine Learning, Deep Learning and digital signal/image processing. His current research interests are object detection, segmentation and medical image analysis.