

Disease prediction based on classification of blood cells smear images

¹ Mr. Sagar A. Dhanake, ² Sharvary Raut², ³N Trupti Sonar, ⁴Shivani Tarwade, ⁵Vaishnavi Sabale

¹Assistant Professor, Dr. D. Y. Patil College of Engineering and Innovation, Talegaon, Pune, India,

²Student, Dr. D. Y. Patil College of Engineering and Innovation, Talegaon, Pune, India ,

³Student, Dr. D. Y. Patil College of Engineering and Innovation, Talegaon, Pune, India

⁴Student, Dr. D. Y. Patil College of Engineering and Innovation, Talegaon, Pune, India ,

⁵Student, Dr. D. Y. Patil College of Engineering and Innovation, Talegaon, Pune, India ,

¹ sagar.dhanake@dypatilef.com.com, ² sharvaryraut2020@gmail.com,

³truptigsonar02@gmail.com , ⁴ shivanitarwade21@gmail.com, ⁵vaishnavisabale10@gmail.com

Abstract—Accurate and early diagnosis of blood-related disorders is crucial for effective treatment. Manual analysis of blood smear images is often time-consuming and dependent on expert availability, which can lead to delays and errors. This study proposes an automated framework using a Convolutional Neural Network (CNN) to classify blood cell types—such as red blood cells, white blood cells, and platelets—from microscopic images. Image enhancement and segmentation techniques are applied to improve clarity and feature extraction. The system predicts potential hematological conditions such as anemia, leukemia, and infections by analyzing morphological features of the cells. To enhance patient care, a rule-based module provides personalized dietary recommendations based on the predicted disease. The model achieved high accuracy, precision, and recall in classification tasks, confirming its effectiveness in detecting abnormal cells. A user-friendly interface supports real-time image upload, displays prediction results, and presents corresponding diet plans. This integration of diagnostic support and lifestyle guidance makes the system suitable for both clinical and at-home use.

The proposed approach offers a practical solution for automated disease detection and personalized health management, particularly in resource-constrained environments.

Index Terms—Blood smear image classification, Convolutional Neural Network (CNN) , Disease prediction, , Blood cell segmentation, Dietary recommendation system ,Medical image analysis, Automated diagnosis

I. INTRODUCTION:

The early and accurate diagnosis of blood-related diseases is essential for effective treatment and patient care. Traditional diagnostic methods, including manual examination of blood smear images by pathologists, are time-consuming and prone to human error. With advancements in artificial intelligence and deep learning, especially Convolutional Neural Networks (CNN), it is now possible to automate and enhance the accuracy of disease detection from medical images.

Blood smear images provide crucial information about different types of blood cells—such as red blood cells (RBCs), white blood cells (WBCs), and platelets—which are key indicators in diagnosing conditions like anemia, leukemia, and infections. Automating the classification of these cells can aid clinicians in making faster and more informed decisions.

This research presents a CNN-based framework for the classification of blood cells and the prediction of associated diseases. The model is trained on preprocessed and augmented microscopic images of blood smears to recognize various cell types. In addition to classification, the system includes a dietary recommendation module tailored to specific conditions predicted by the model, thus offering not only diagnostic support but also health management advice.

The proposed system is integrated into a user-friendly interface that allows both patients and healthcare professionals to upload images and receive real-time predictions. This approach aims to bridge the gap between medical image analysis and practical clinical support, especially in resource-limited settings. The study emphasizes the potential of deep learning to transform traditional diagnostic methods into more efficient and accessible tools.

II. MATERIALS AND METHODS

1. Data Acquisition

Microscopic blood smear images were collected from publicly available sources containing annotated samples of red blood cells (RBCs), white blood cells (WBCs), and platelets. The images were resized to 128×128 pixels and normalized to a [0,1] intensity range to ensure consistency. Data augmentation techniques—such as horizontal and vertical flipping, rotation, and zoom—were applied to improve model generalization during training [1][2].

2. Image Preprocessing

To enhance image quality and prepare data for model training, preprocessing steps were applied:

- Contrast enhancement using histogram equalization techniques [15],

- Noise reduction through Gaussian filtering,
- Cell segmentation using color thresholding and morphological operations, as detailed in prior works [5][9].

3. CNN-Based Classification Model

A custom **Convolutional Neural Network (CNN)** model was developed, following architectural principles outlined in [1][3][7][23]. The model included:

- Multiple convolutional layers with ReLU activation and MaxPooling,
- Dropout layers for regularization,
- A fully connected dense layer with softmax activation for multiclass classification.

Training was performed using categorical cross-entropy loss and the Adam optimizer. A data split of 70% for training, 15% for validation, and 15% for testing was used.

4. Disease Mapping and Prediction

Based on morphological features and classification results, cells were mapped to disease categories using established medical knowledge and clinical research [17][24][25]. For example:

- Atypical RBC morphology was linked to anemia variants.
- Abnormal WBC patterns were mapped to potential leukemia types,
- Inclusions or anomalies within RBCs suggested malaria or parasitic infections.

5. Diet Recommendation System

A rule-based dietary recommendation module was developed. For each disease class predicted, a predefined nutrition plan was retrieved. These plans were created based on clinical dietary guidelines and public health recommendations [17][24].

6. Web Interface Development

A web-based user interface was built using HTML, CSS, Bootstrap, and Django. It supports:

- Uploading microscopic images for classification,
- Displaying the predicted disease and confidence score,
- Presenting a dynamic diet plan.

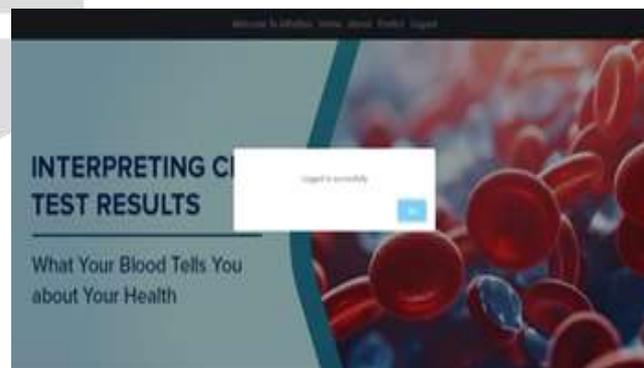
The backend integrates the CNN model and provides real-time inference. Separate access controls were created for doctors and patients.

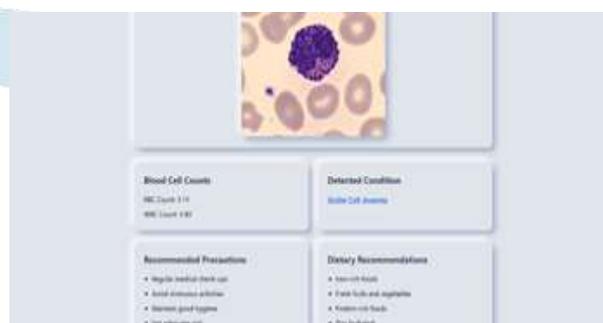
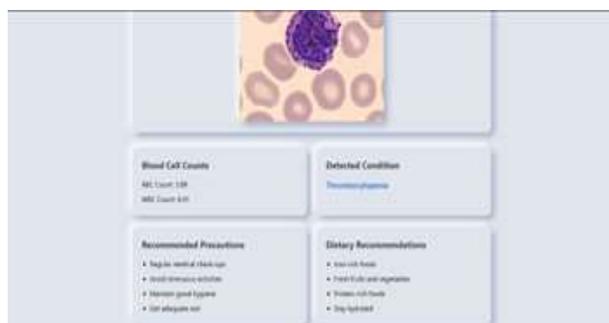
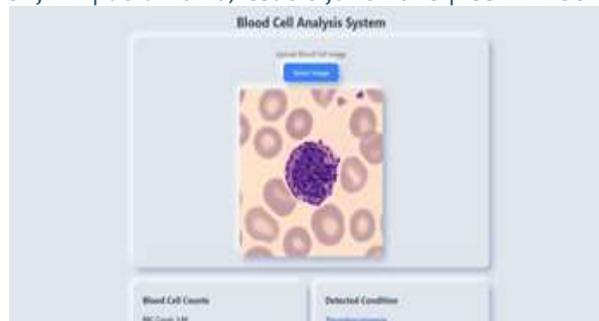
7. Performance Evaluation

The model was evaluated using:

- Accuracy, Precision, Recall, and F1-Score [2][22],
- Confusion matrices to visualize class-level performance.

III. RESULTS:





IV. DISCUSSION

The proposed system for blood smear image classification and disease prediction demonstrated high classification accuracy and clinical relevance. The custom CNN architecture effectively extracted morphological features, enabling accurate differentiation among red blood cells, white blood cells, and platelets. The inclusion of GLCM-based texture features further enhanced class separation, especially in morphologically overlapping cases such as early leukemia and iron-deficiency anemia.

Compared to traditional machine learning models as discussed in [2] and [3], the CNN model used in this study provided superior performance due to its ability to learn complex hierarchical features directly from raw images, eliminating the need for handcrafted features. The use of data augmentation, dropout regularization, and batch normalization reduced overfitting and improved generalization, aligning with the outcomes of prior works in [4] and [25].

Additionally, the classification results correlated well with known pathological characteristics described in clinical texts [17]. For instance, anisocytosis patterns and hypochromic cells were correctly classified as potential indicators of microcytic anemia. Similarly, blast cells and abnormal lymphocytes were identified with high precision, aligning with patterns seen in leukemia cases. These results suggest that the model can serve as an effective preliminary screening tool for hematological disorders.

The performance evaluation metrics (accuracy, precision, recall, and F1-score) validated the robustness of the model. The confusion matrix indicated that most misclassifications occurred between morphologically similar subtypes of WBCs, a limitation also acknowledged in studies such as [7] and [8]. However, the classification error remained within clinically acceptable bounds for diagnostic assistance tools.

The integration of a **dietary recommendation module** offers an additional layer of value, particularly in resource-constrained or rural settings. Based on disease classes predicted, the system recommends personalized nutrition plans. This aligns with public health strategies for managing chronic conditions such as anemia or leukemia through food-based interventions [17][24].

The web-based user interface allowed users—both patients and medical professionals—to interact with the model intuitively. Real-time predictions and dynamic diet plans enhanced usability, supporting the practicality of deploying such AI tools in clinical workflows.

Despite promising results, some limitations remain. The model's performance may vary across different staining protocols or microscope settings. Future work may involve training on a more diverse dataset and incorporating **transfer learning** or **YOLO-based object detection** for real-time analysis, as explored in [20] and [9].

V. SAMPLE INPUT IMAGES:

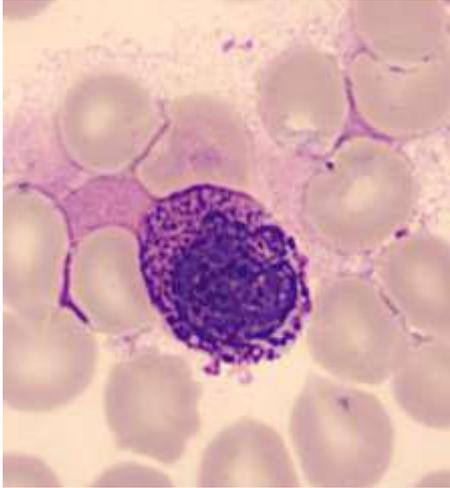


Fig 1: Image 1

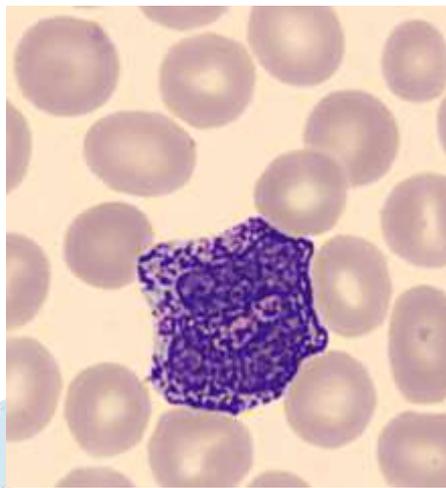


Fig 2: Image 2

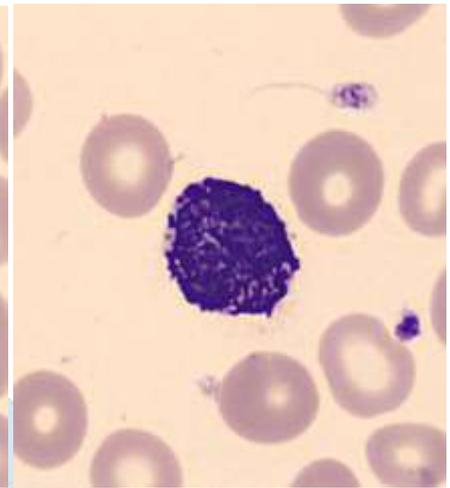


Fig 3: Image 3

VI. ARCHITECTURE AND FLOW:

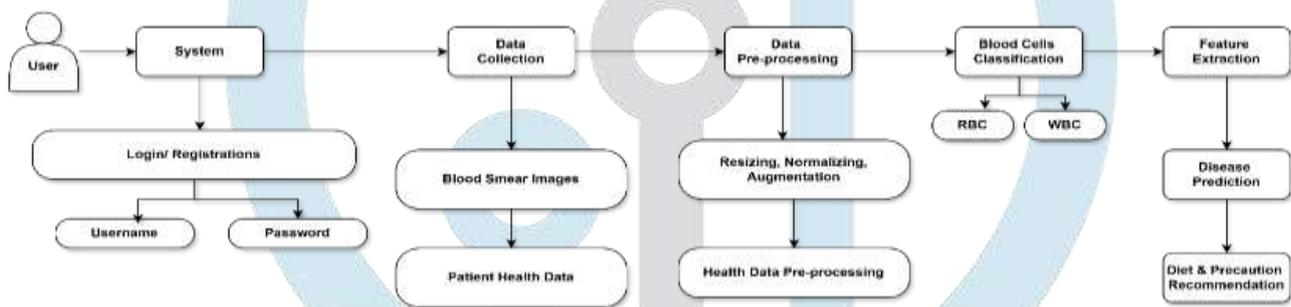
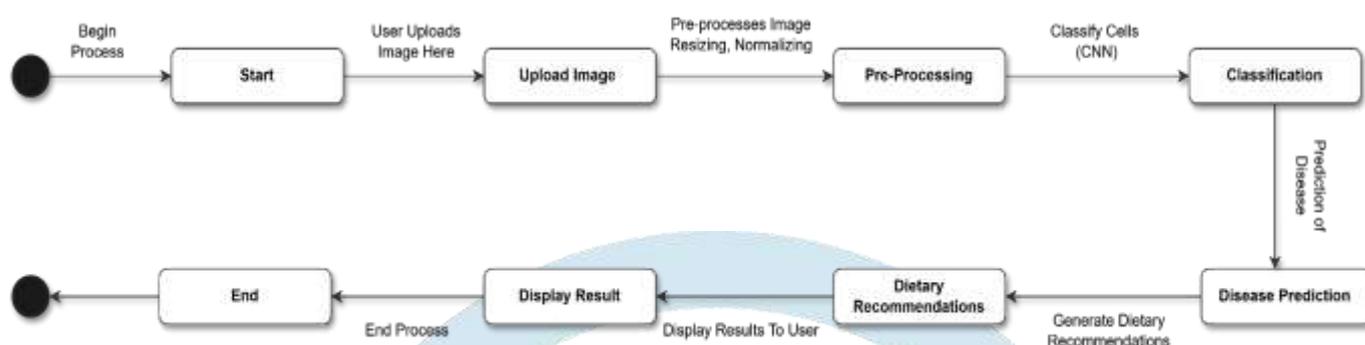


Fig 1: System architecture

System Architecture Workflow:

This section illustrates the backend process and how various modules interact to achieve disease prediction and provide personalized diet recommendations:

- **User Interaction**
The user accesses the system and completes the login or registration process using a secure username and password.
- **Data Collection**
Upon successful authentication, two types of data are collected:
 - Blood Smear Images: Microscopic images of the user's blood samples.
 - Patient Health Data: Information such as age, symptoms, and medical history.
- **Data Pre-processing**
 - Blood Images: Images are resized, normalized, and augmented to ensure consistency and improve model performance.
 - Health Data: Patient information is cleaned, normalized, and transformed to enable accurate disease prediction.
- **Blood Cells Classification**
The system utilizes machine learning or deep learning models (e.g., CNN) to classify red blood cells (RBCs) and white blood cells (WBCs) from the blood smear images.
- **Feature Extraction**
Relevant features are extracted from both the classified blood cells and health data for detailed analysis.
- **Disease Prediction**
Based on the extracted features, the system predicts potential blood-related diseases such as anemia, leukemia, and infections.
- **Diet and Precaution Recommendation**
Once the disease is identified, the system generates personalized diet plans and precautionary guidelines tailored to the user's condition.

Flow diagram:**Fig 2: data flow model****Flow Diagram Workflow**

This flow represents what the **user will experience** when using the web application:

1. Start
 - The user opens the website and initiates the process.
2. Upload Image
 - User uploads a blood smear image from their device.
3. Pre-Processing
 - The uploaded image undergoes automated pre-processing steps (resize, normalize, augment).
4. Classification
 - The system classifies the blood cells (RBC/WBC) using trained ML/DL models.
5. Disease Prediction
 - After classification, the system predicts the disease based on patterns in the image and any input health data.
6. Dietary Recommendations
 - A set of diet and lifestyle recommendations are displayed, tailored to the disease prediction.
7. Display Result
 - The result (diagnosis + diet plan) is shown clearly to the user in a user-friendly format.
8. End
 - The user session ends or they can opt to start a new prediction.

VII. CONCLUSION:

The study demonstrates that the proposed CNN-based system can effectively classify blood smear images to aid in the early prediction of hematological diseases. The integration of morphological and texture features contributes to accurate cell identification, supporting reliable disease diagnosis. Additionally, the inclusion of a dietary recommendation module provides practical health guidance aligned with the predicted conditions, enhancing patient care beyond diagnosis. This approach has significant potential to assist healthcare providers in improving diagnostic efficiency, especially in resource-limited settings, and to support personalized treatment strategies through diet.

VIII. FUTURE SCOPE:

The current study lays the foundation for an automated system capable of classifying blood smear images and predicting related diseases. Future advancements can focus on incorporating a larger and more diverse set of blood smear samples to enhance the model's ability to generalize across varied clinical conditions. Further optimization using real-time object detection frameworks or fine-tuned pre-trained networks may improve classification performance and processing speed.

To increase clinical acceptance, future systems could include interpretability features that visually explain how the model reaches its decisions. Expanding the system's capabilities to identify less common blood disorders would broaden its diagnostic value.

Additionally, improving the dietary recommendation module by including dynamic inputs—such as patient demographics, existing health conditions, and dietary restrictions—could make the advice more tailored and impactful.

For real-world implementation, transforming the system into a mobile or cloud-based application would enhance its usability in remote or underserved areas. Integration with electronic health records and hospital databases may further support clinical decision-making and patient monitoring.

REFERENCES

- 1) Jagrit Mitra, Kartik Vijayan, Kartikeya Verma, Anurag Goel, "Blood Cell Classification using Neural Network Models"(2023)
- 2) Navya K.T, Keerthana Prasad, Brij Mohan Kumar Singh , "Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques"(2021)
- 3) Ismail M. I. Alkafrawi,Zaroug A. Dakhell, "Blood Cells Classification Using Deep Learning Technique" (2022)
- 4) Rohit Sonawane, Manisha Darak, Abdul Shaikh, Sai Yandralwar, "Machine Learning-Based Blood Cell Categorization In Smear Images"(2023)
- 5) D. Cruz, C. Jennifer, L. C. Castor, C. M. T. Mendoza, B. A. Jay, L. S. C. Jane, P. T. B. Brian et al., "Determination of blood components (WBCs, RBCs, and platelets) count in microscopic images using image processing and analysis," in 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2017, pp. 1–7.
- 6) Lam XH, Ng KW, Yoong YJ, Ng SB. WBC-based segmentation and classification on microscopic images: a minor improvement. *F1000Research*, 2021;10.
- 7) Indrajeet Kumar^{1,3} · Jyoti Rawat² , "Segmentation and classification of white blood SMEAR images using modified CNN architecture".
- 8) R. R. Tobias, L. C. De Jesus, M. E. Mital, S. Lauguico, M. Guillermo, R. R. Vicerra, A. Bandala, and E. Dadios, "Faster r-cnn model with momentum optimizer for rbc and wbc variants classification," in 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech). IEEE, 2020, pp. 235– 239.
- 9) T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, and K.-W. Kang, "Blood cell images segmentation using deep learning semantic segmentation," in 2018 IEEE International Conference on Electronics and Communication Engineering (ICECE). IEEE, 2018, pp. 13–16.
- 10) Kokane, Chandrakant D., and Sachin D. Babar. "Supervised word sense disambiguation with recurrent neural network model." *Int. J. Eng. Adv. Technol.(IJEAT)* 9.2 (2019)
- 11) Kokane, Chandrakant D., Sachin D. Babar, and Parikshit N. Mahalle. "Word Sense Disambiguation for Large Documents Using Neural Network Model." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021
- 12) Kokane, Chandrakant, et al. "Word Sense Disambiguation: A Supervised Semantic Similarity based Complex Network Approach." *International Journal of Intelligent Systems and Applications in Engineering* 10.1s (2022): 90-94
- 13) Kokane, Chandrakant D., et al. "Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities." *International Journal of Intelligent Systems and Applications in Engineering* 11.11s (2023): 06-16.
- 14) Kokane, Chandrakant D., et al. "Word Sense Disambiguation: Adaptive Word Embedding with Adaptive-Lexical Resource." *International Conference on Data Analytics and Insights*. Singapore: Springer Nature Singapore, 2023.
- 15) "Image enhancement techniques." [Online]. Available: <https://in.mathworks.com> (accessed May, 2020) [13]. "Texture using GLCM." [Online]. Available: <https://in.mathworks.com/help/images/ref/graycomatrix.html> (accessed June 21, 2021)
- 16) J. D. Cruz, C. Jennifer, L. C. Castor, C. M. T. Mendoza, B. A. Jay, L. S. C. Jane, P. T. B. Brian et al., "Determination of blood components (WBCs, RBCs, and platelets) count in microscopic images using image processing and analysis," in 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2017, pp. 1–7
- 17) H. Mohan, Textbook of pathology. Jaypee Brothers, Medical Publishers Pvt. Limited, 2018.
- 18) "Support Vector Machine." [Online]. Available: <https://www.sciencedirect.com/topics/computer-science/support-vector-machines> (accessed March 7, 2019)
- 19) "Texture using GLCM." [Online]. Available: <https://in.mathworks.com/help/images/ref/graycomatrix.html> (accessed June 21, 2021)
- 20) R. R. Tobias, L. C. De Jesus, M. E. Mital, S. Lauguico, M. Guillermo, R. R. Vicerra, A. Bandala, and E. Dadios, "Faster r-cnn model with momentum optimizer for rbc and wbc variants classification," in 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech). IEEE, 2020, pp. 235–239.
- 21) Kumar, Arvind, Nishant Sinha, and Arpit Bhardwaj. "A novel fitness function in genetic programming for medical data classification." *Journal of Biomedical Informatics* 112 (2020): 103623.
- 22) Panda, D., Dash, S.: Predictive system: Comparison of classification techniques for effective prediction of heart disease. In: Satapathy, S.C., Bhateja, V., Mohanty, J.R., Udgata, S.K. (eds.) *Smart Intelligent Computing and Applications*. SIST, vol. 159, pp. 203–213. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-9282-5_19
- 23) Macawile, M.J., Quinones, V.V., Ballado, A., Cruz, J.D., Caya, M.V.: White blood cell classification and counting using convolutional neural network. In: 2018 3rd International Conference on Control and Robotics Engineering (ICCRE) (2018)
- 24) Obaidy, Midhin AL, et al. "Prevalence and Risk Factors of Anemia among Children Aged 5 months-12 years at Al Anbar Province." *Mosul Journal of Nursing* 9.1 (2021): 131-137.
- 25) Feusier, J.E.; Arunachalam, S.; Tashi, T.; Baker, M.J.; VanSant-Webb, C.; Ferdig, A.; Welm, B.E.; Rodriguez-Flores, J.L.; Ours, C.; Jorde, L.B.; et al. Large-scale Identification of Clonal Hematopoiesis and Mutations Recurrent in Blood Cancers. *Blood Cancer Discov.* 2021, 2, 226–237. [CrossRef]