



NORTH SOUTH UNIVERSITY  
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING  
**CSE499A.5: SENIOR DESIGN**

# **FINAL PROJECT PRESENTATION SLIDE**

**Submitted To:**

**Dr. Mohammad Moniruzzaman Khan**

**Submitted By:**

**Adit Hossain- 1813612642**

**Moontasir Fuad Plabon-1921541042**

**Nafisa Yasmin- 2011771642**

# Topic

**Brain Cancer Gene Expression Detection Using Machine Learning**

# Background and history about brain cancer

- In recent times, cancer has become more fluent. It is one of the leading causes of mortality in the entire world.
- Brain cancers include primary brain tumors, which start in the brain and rarely spread to other parts of the body, and secondary tumors (or metastases), which are caused by cancers that began in another part of the body.
- Primary brain cancer begins when normal cells develop changes (mutations) in their DNA. A cell's DNA contains the instructions that tell a cell what to do. The mutations tell the cells to grow and divide rapidly and to continue living when healthy cells would die.
- The outlook for a malignant brain tumor depends on where it is in the brain, its size, and what grade it is, which turns into brain cancer later. It can sometimes be cured if caught early on, but a brain tumor often comes back, and sometimes it isn't possible to remove it at all.

# PROBLEM STATEMENT

- People nowadays often have cancer but can't tell it. Brain cancer is the most dangerous cancer of all.
- Due to this, sometimes, when they receive treatment from a doctor, they can not tell the doctor about brain cancer. So they can not get proper treatment for it, and the chance that something serious can occur increases.
- Sometimes, it also takes a little while for the doctor to learn about this type of cancer gene. So, using ML, we can detect brain cancer genes easily.

# SIMILAR SYSTEM REVIEW

Medical diagnostics and preventative medicine are just two areas where machine learning techniques have been heavily applied. However, only a small number of studies have focused specifically on brain tumor diagnosis. Most ML techniques use MRI data to test and train conventional ML algorithms. Recently, some methodologies have used DL to diagnose brain tumors. Rehman et al. [5] established a framework for classifying cancers of various sorts using a setup known as tri-architectural CNN (convolution neural network) (GoogLeNet, AlexNet and VGGNet). This categorization included meningioma, glioma, and pituitary tumor varieties. The aforementioned algorithm divided the brain MRI into interested parts. The data sets were also fine-tuned and frozen in preparation for additional classification. To ensure the veracity of the results, the authors also took data augmentation approaches into account. With the help of the VGG16 architecture, this study's classification and detection accuracy increased to 92.69%.

# **Objective and purpose of this project**

- The main objective of this project is to get the highest possible prediction accuracy for the data used here.
- To help people detect if they have cancer or not, and to make people aware of that if there is a chance cancer could occur.
- To make a reputable and trustworthy product that can predict brain cancer with a higher accuracy rate while using the input dataset.

# POSSIBLE SOLUTION

A DNA microarray can monitor thousands of genes' levels of expression at once. Previous studies have shown that this technology can help in cancer classification. Typically, cancer microarray data comes from a limited number of samples with a significant number of features related to gene expression levels. In this study, feature selection algorithms were extensively studied to minimize dimensionality and extract relevant gene information from cancer microarray data. We demonstrate that classification performance on acute leukemia and diffuse large B-cell lymphoma microarray data sets can be obtained using a correlation-based feature selector in conjunction with machine learning algorithms such as decision trees, naive Bayes, and support vector machines. We also show that it is possible to choose relevant genes with a high degree of confidence by combining the use of various categorization and feature selection methodologies.

There are two main types of strategies: filters and wrappers. The majority of filter methods assess features by rating them in accordance with the overall traits of the training set. Then, they eliminate characteristics that are unnecessary by setting a threshold. A gene will be chosen if its score is higher than the cutoff. Additionally, some filtering techniques, like CFS, give a score to subsets of features. Wrapper methods, in contrast, consider the biases of machine learning algorithms when choosing features. They employ cross-validation to calculate a score for feature subsets after applying a machine learning algorithm to them.

# Output of the Project

The primary goal is to present the state-of-the-art in the field of brain cancer, which includes the pathophysiology of the disease, imaging technologies, WHO classification standards for tumors, primary methods of diagnosis, and existing computer-assisted algorithms for brain cancer classifications using machine and deep learning techniques. We concluded by contrasting brain tumors with other conditions of the brain. We have come to the conclusion that, in comparison to traditional classification techniques or medical imaging, DL-based methods are currently receiving more attention and accuracy due to their automatic feature extraction capacity. There is little doubt that many lives can be spared if cancer is found and its appropriate grade is determined using quick and affordable diagnosis methods. Therefore, the development of quick, non-invasive, and affordable diagnosis tools is imperative. Here, DL techniques may be very important for the same. As far as we are aware, relatively little research has been done on automatically grading tumors utilizing DL approaches, and their full potential has not yet been discovered.

# **ALGORITHM USED**

- 1.** Support vector machine recursive feature elimination (SVM-RFE)
- 2.** Twin support vector machine (TWSVM)
- 3.** Support Vector Machine (SVM)
- 4.** K-Nearest Neighbor (KNN)
- 5.** Naive Bayes Algorithm

# **SOFTWARE USED**

- 1. TensorFlow**
- 2. Anaconda**
- 3. Google Colab**
- 4. Visual Studio Code**

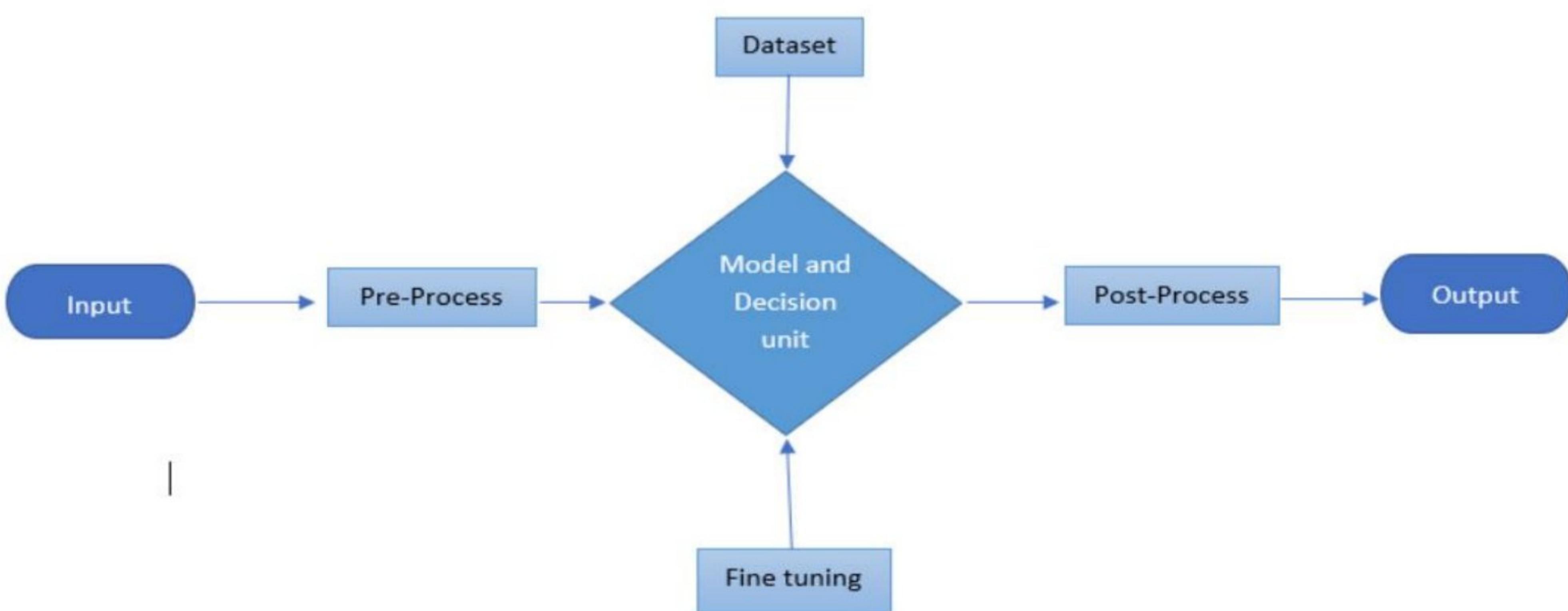
# **Programming Language Used**

**Python**

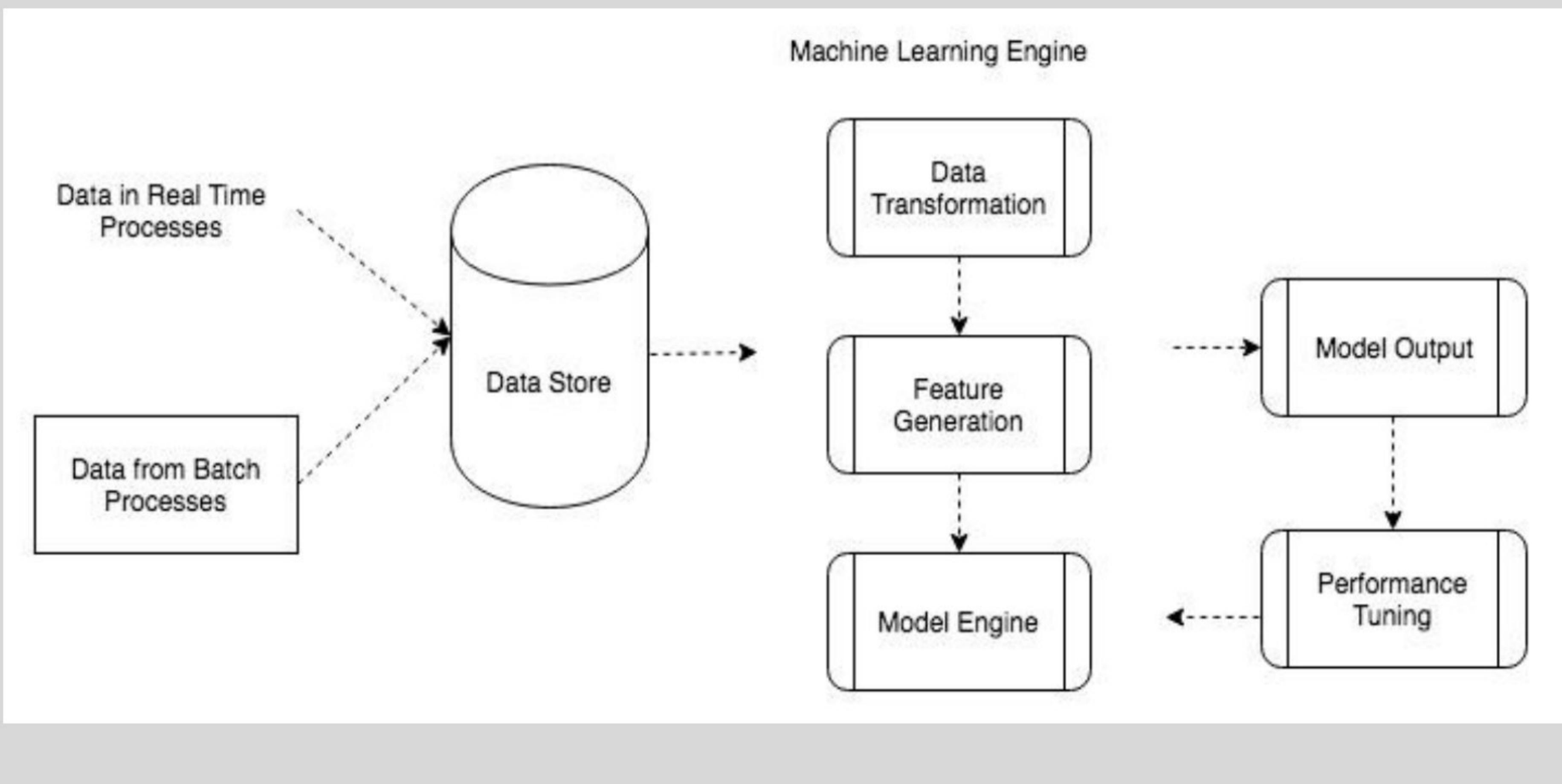
# Dataset Used

B118		X	✓	f <sub>x</sub>	pilocytic_astrocytoma																											
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	
2	834	ependyr	12.498	7.6049	6.8809	9.0271	4.1762	7.2249	6.0859	6.836	5.8984	5.5134	7.542	7.9051	7.1081	8.3866	8.4121	6.5601	5.7793	5.5961	9.1289	6.8228	8.2223	5.9657	6.0708	6.5227	7.0314	6.5148	8.3335	6.6999	8.2906	6.
3	835	ependyr	13.067	7.9981	7.2091	9.7233	4.8261	7.5394	6.251	8.0125	5.4531	6.1731	8.317	8.2691	7.2507	9.2086	8.4662	5.822	5.8629	5.5732	8.8535	5.8559	6.9146	6.857	7.2684	7.8058	7.9536	7.0492	7.9859	8.2852	9.7595	5.
4	836	ependyr	13.068	8.5737	8.6477	9.613	4.3966	7.8131	6.0077	7.1782	8.4003	6.3235	7.6619	8.3279	7.5572	9.3446	9.1613	6.3525	5.704	6.2969	8.9123	6.1153	6.7065	7.2362	7.0628	7.7028	7.7137	7.1466	8.9583	5.9666	8.2356	5.
5	837	ependyr	12.456	9.099	6.6288	8.5177	4.1548	8.3618	6.5961	6.3473	4.9004	6.0087	7.6638	7.7302	6.8909	8.9964	7.9677	5.5654	5.6569	7.2554	7.7259	5.8355	8.9859	6.6888	6.5825	7.0736	7.6032	6.4205	9.4777	5.9219	7.7753	4.
6	838	ependyr	12.7	8.8007	11.556	9.1663	4.1659	7.9238	6.2128	6.8664	5.4056	5.2796	9.522	7.5814	6.9276	10.041	9.6114	5.9932	5.8744	6.6358	9.7218	5.6033	8.2218	5.9504	5.7528	7.9164	8.1398	6.7399	9.6203	6.8691	8.2656	5.
7	839	ependyr	12.461	7.6765	7.2445	9.6769	3.9043	7.4671	6.5852	7.0616	5.2848	5.1773	7.2811	8.031	7.2997	8.6736	8.8686	5.9531	5.7332	5.9407	9.5923	7.9817	8.7042	6.49	6.8986	7.3109	6.8843	6.5129	8.6712	6.8095	8.4097	5.
8	840	ependyr	13.656	7.9809	7.5666	10.103	4.6594	8.4577	6.1874	7.5334	6.5559	5.6399	8.3571	8.4398	7.7725	9.8804	8.8453	6.2453	5.9984	6.5972	9.304	5.9134	8.7701	7.3031	7.1189	7.6057	7.8123	7.0509	8.2642	8.1976	9.514	5.
9	841	ependyr	12.601	8.8732	9.2143	9.6903	4.3418	7.9246	6.4826	6.4996	5.369	6.6348	7.434	7.8971	7.1158	8.8261	8.9524	6.0856	6.8134	6.2762	8.5612	5.9554	8.3665	6.7643	6.6935	7.2276	6.7171	6.885	8.5658	6.8061	8.3937	5.
10	842	ependyr	12.463	8.2065	8.8912	9.7373	3.9363	7.9063	6.0662	6.9809	6.4266	6.0358	7.4257	8.1015	6.9718	8.8542	8.9954	6.1444	5.9503	6.0981	8.5025	6.4384	6.4407	6.2136	5.9746	7.8859	7.4172	6.9165	8.8748	6.3683	8.2671	5.
11	843	ependyr	12.266	7.8799	10.211	9.701	4.0877	8.102	6.5104	6.9866	6.4019	5.3063	7.6593	8.2196	7.5616	8.8324	8.3159	6.2678	6.2345	6.2415	8.9173	6.5009	8.3413	6.7444	6.5194	7.7087	7.5098	7.0428	9.1979	6.4559	8.4766	5.
12	844	ependyr	12.921	8.6205	7.2386	8.2454	3.6828	8.0274	7.5753	5.5175	6.7612	5.1802	6.3559	8.7629	6.7033	9.6805	9.287	5.4253	5.5049	7.1576	8.0352	5.4153	7.4876	6.7837	6.5054	7.5962	7.5673	6.0645	9.2083	6.6234	7.7226	4.
13	845	ependyr	12.787	7.5708	7.519	9.2976	4.0409	8.7794	6.6266	7.3084	5.1291	5.6498	9.9961	9.2535	6.6589	10.627	7.627	6.3873	6.1887	5.6731	8.7465	6.2777	6.5057	6.3362	5.7842	8.075	8.579	7.0192	7.1056	8.506	9.5234	5.
14	846	ependyr	12.545	8.4135	7.2506	10.101	4.2963	8.1972	6.4718	7.3436	4.702	5.259	7.8633	7.8587	7.1499	8.6172	8.6085	6.0639	6.0516	5.5708	9.2143	5.8906	8.4899	6.9782	6.5493	6.4648	6.7016	7.219	8.3743	6.6847	8.7915	5.
15	847	ependyr	12.588	9.3813	7.1646	8.5457	6.1785	7.3391	6.3751	5.9799	5.5808	5.2379	9.4632	8.7559	6.4339	10.473	9.5077	5.8543	5.1943	6.9688	8.7183	5.2353	6.3528	7.0277	7.2777	7.7937	7.7641	6.7889	8.8556	7.3118	8.7063	4.
16	848	ependyr	13.235	8.7912	7.3969	9.6201	4.0423	8.1879	6.2092	7.855	5.8683	5.4678	8.3057	7.5687	7.0248	9.1261	9.3065	5.7331	5.6572	6.5935	9.7782	7.0047	8.8229	7.0021	7.1116	7.6452	7.8023	6.9409	8.873	6.398	8.5054	5.
17	849	ependyr	12.699	9.1602	12.054	8.891	4.2433	7.8851	6.3001	6.7527	5.4004	4.9018	7.7885	7.7595	7.0215	9.8029	9.2144	5.5064	5.634	6.1644	8.6483	6.3455	10.12	6.7277	6.2599	7.4428	7.4955	6.4971	10.221	6.9056	8.5413	4.
18	850	ependyr	13.53	7.6211	7.8935	10.176	4.3593	8.6924	6.5686	7.9333	6.0182	5.6262	8.6459	8.1576	7.6639	10.385	8.9148	6.0986	6.1812	6.771	9.7178	5.9044	8.1793	7.3768	7.0754	7.4105	7.7348	7.3599	8.3762	8.5292	9.6447	5.
19	851	ependyr	13.268	8.1162	8.3193	9.7102	4.0938	7.8151	6.683	6.6382	5.5858	6.0029	7.819	8.8313	7.4804	9.0756	9.3453	5.9434	5.7031	6.2265	8.4355	6.0943	7.0938	7.2822	6.8287	7.2428	6.9848	6.7225	8.7751	6.2969	8.4737	5.
20	852	ependyr	12.884	8.9805	8.2629	8.68	4.2265	8.3813	6.5383	7.2866	5.0584	5.7281	7.3837	7.6704	6.654	9.2177	9.4677	6.0487	5.6225	6.8392	9.5828	5.5739	8.2875	6.9809	6.57	7.6728	7.8501	6.644	9.3458	6.5332	7.9213	4.
21	853	ependyr	13.142	8.6728	7.1201	9.5308	4.2671	7.9707	6.552	6.9834	6.2725	5.2425	7.4744	8.0622	7.0617	8.9951	9.1662	6.1612	5.612	6.085	9.6304	6.5979	9.62	7.1059	7.0835	7.2481	7.4829	7.3019	8.983	7.0996	8.3029	5.
22	854	ependyr	13.126	8.7769	8.9047	9.7251	4.45	7.619																								

# BLOCK DIAGRAM



# System Architecture



# GANNT CHART

SL	Milestones	Month 1	Month 2	Month 3	Month 4	Month 5	Total Time Period
1	Topic Selection	.....					1 month
2	Background Study		.....				1 month
3	Software Design			.....	.....		2 months
4	Testing				.....		1 month
5	Analysis					.....	1 month

# TARGET POPULATION

Every citizen is the target group. It is very depressing to observe how many elderly individuals, particularly in urban settings, struggle to obtain adequate medical care. However, the problem can be successfully solved with this project. Senior adults can receive medical care from doctors and hospitals without having to leave their homes. They can simplify and improve their quality of life with the aid of technology. This will benefit those with impairments as well as individuals in general, especially seniors, who will benefit from an improvement in their health.

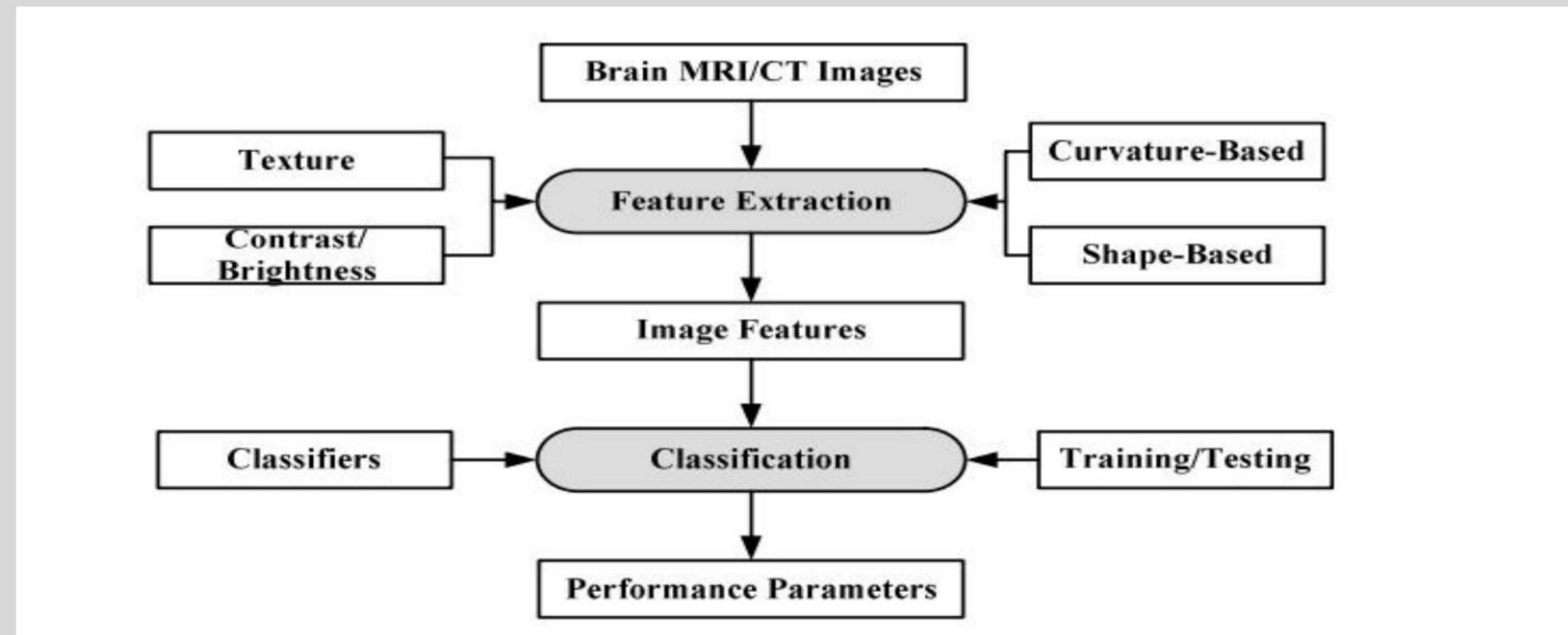
## BENEFIT FROM THE PROJECT

This will help to predict whether they have cancerous cell in their body or not by 98% accuracy and will help them in so many ways. This will save a lot of time and money for the patient and as well as for the hospital. With the help of datasets and image one can predict whether they have cancerous cells or not.

# **Unprivileged women and people will benefit?**

Yes it does help the under privileged people. All kind of people can be benefited from this project.

# THE SYSTEM FUNCTIONING



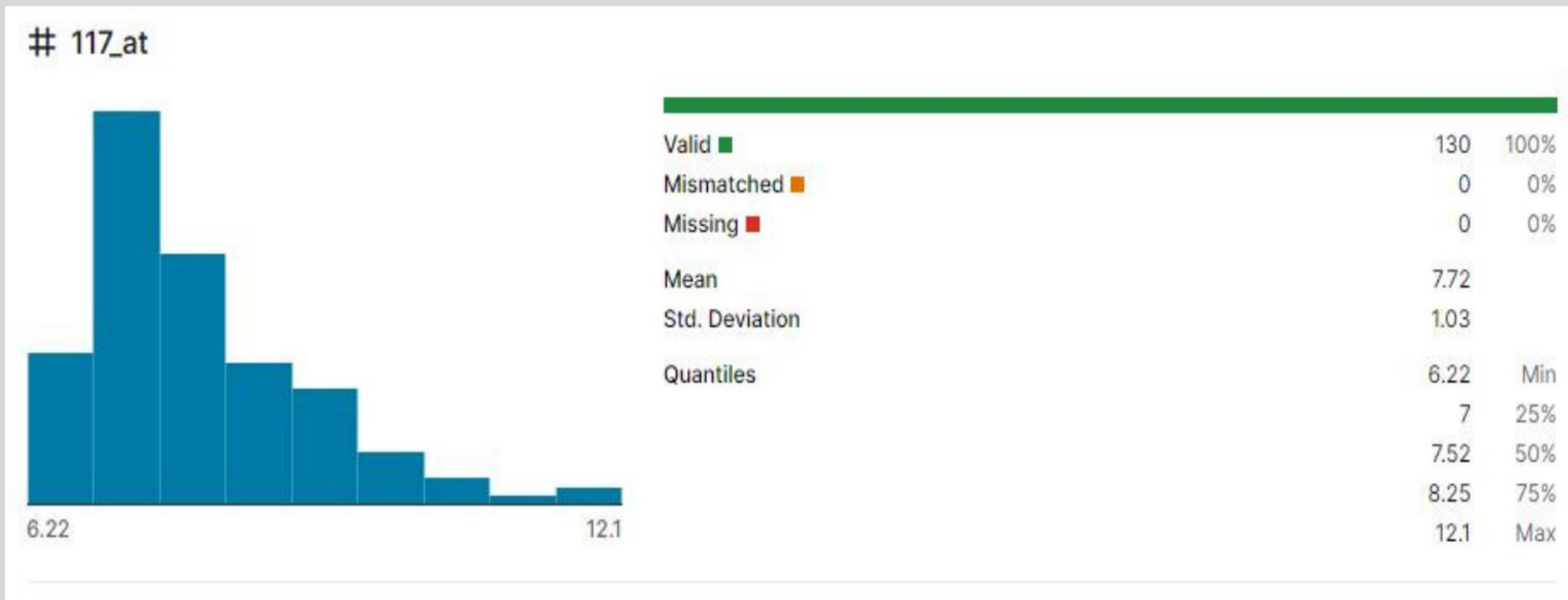
# Some Details About Our Project

Classification of brain cancer is a problem of multiclass classification. One approach to solving this problem is transforming it into several binary problems. The dataset we will be using has four types of prediction: pilocytic astrocytoma, medulloblastoma, ependymoma, or normal. It is a classification type of problem which have four classes in it. Another one is the microarray gene expression dataset which has two main characteristics of medical data: extremely many features (genes) and only a few numbers of samples. The application of machine learning on microarray gene expression datasets mainly consists of two steps: feature selection and classification. The paper we read selects features based on the support vector machine recursive feature elimination (SVM-RFE) principle, which is improved to solve multiclass classification, called multiple multiclass SVM-RFE.

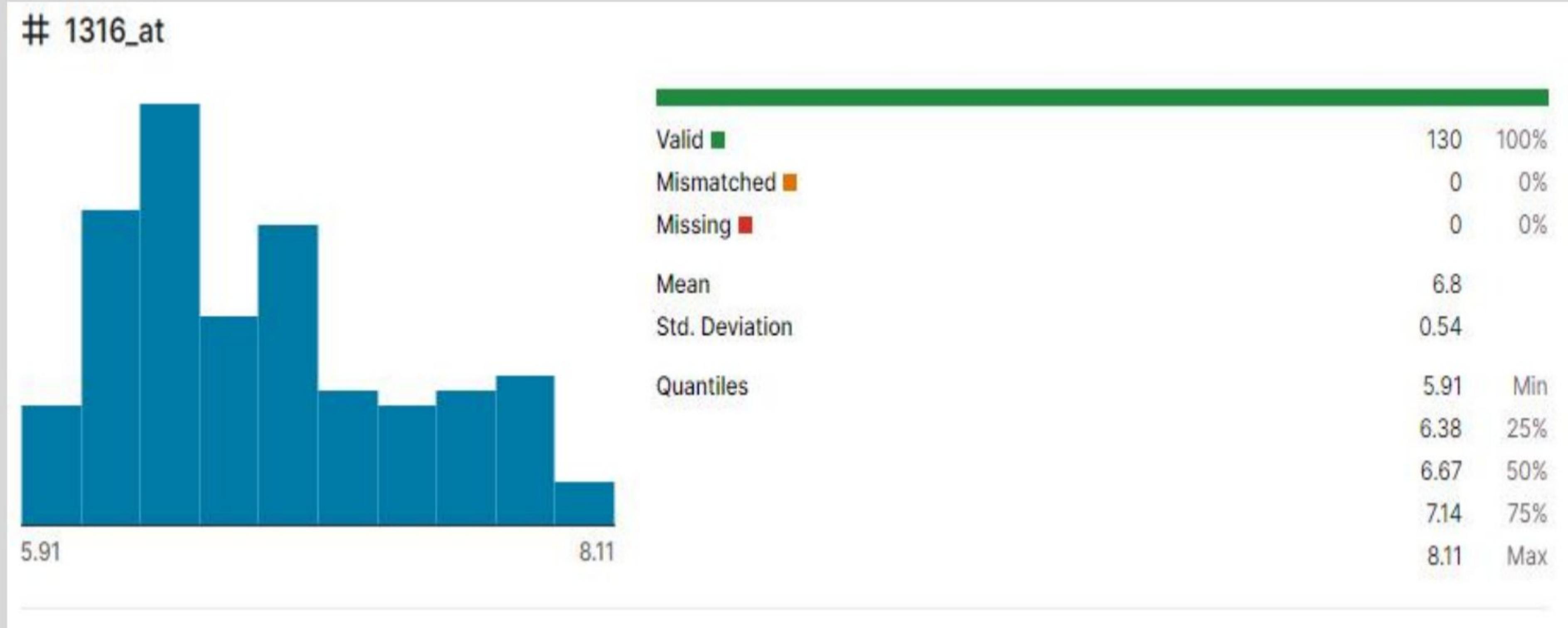


# RESULTS

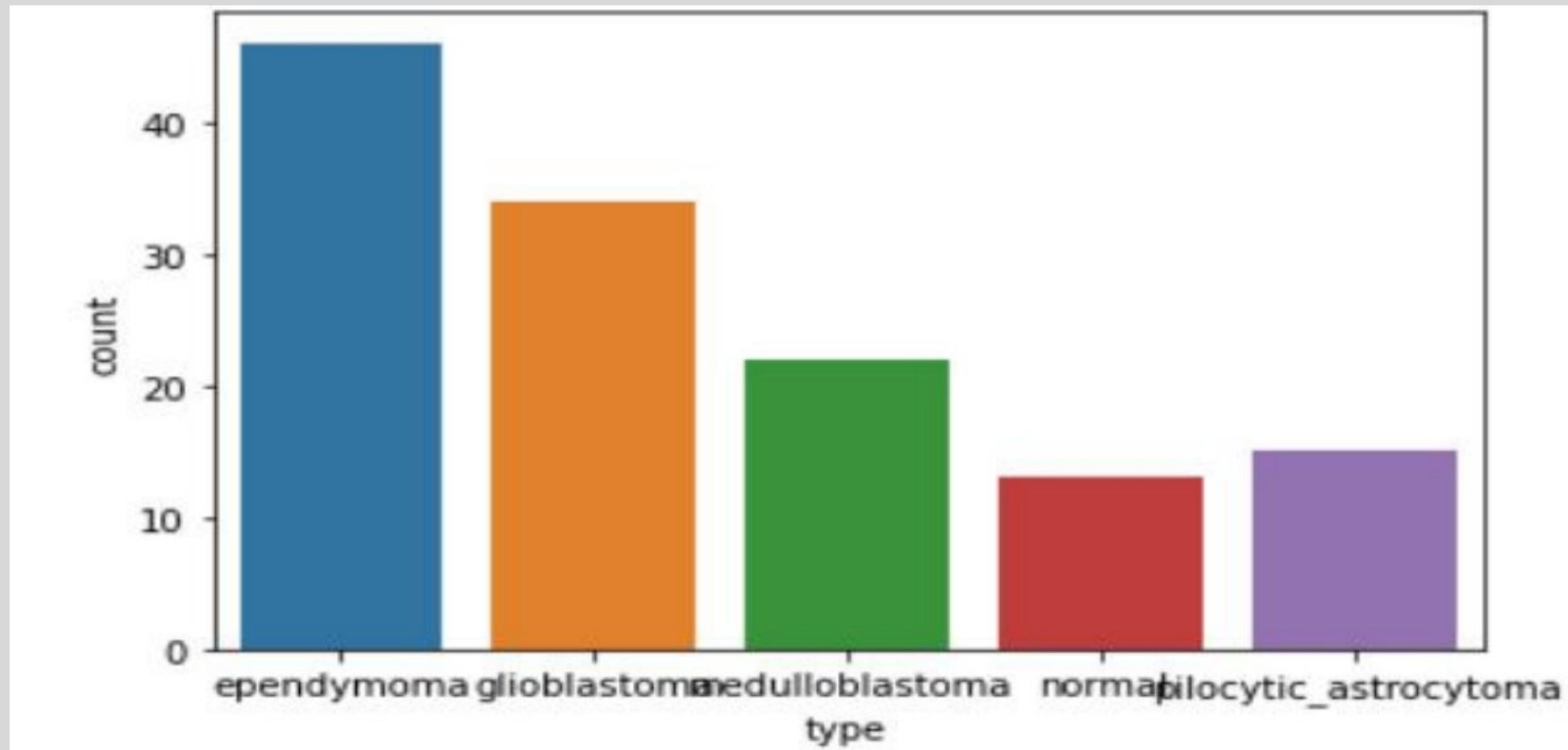
# TABLE SHOWING INPUT AND OUTPUT DATA



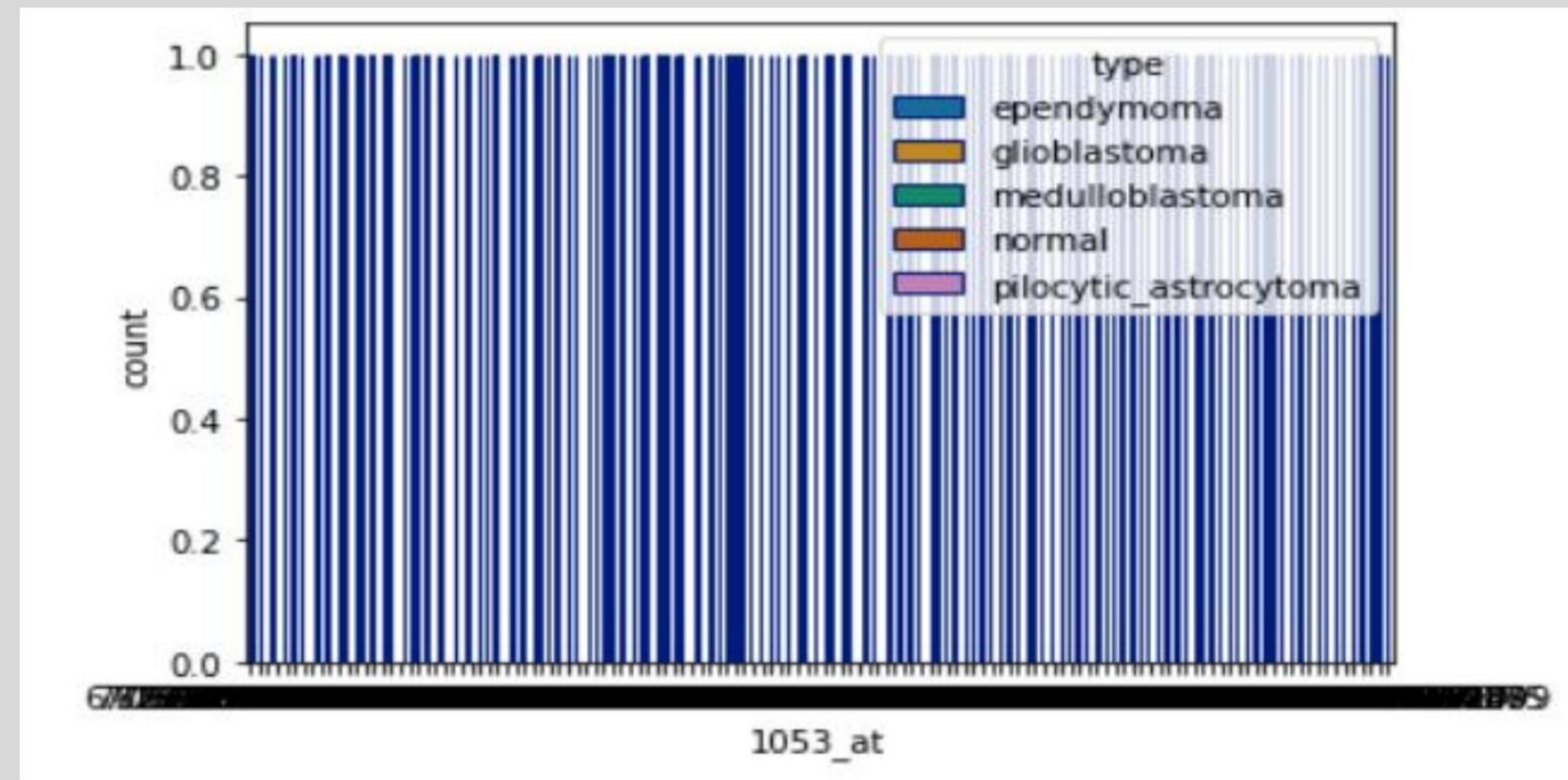
# Data Preprocessing



# Types of output in Dataset



# Figure showing count from type variable



# Training accuracy

```
#again some more library for the training
from numpy import mean
from numpy import std
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedStratifiedKFold

cv = RepeatedStratifiedKFold(n_splits=5, n_repeats=3, random_state=1)
n_scores = cross_val_score(model, X, y, scoring='accuracy', cv=cv, n_jobs=-1)
print('Mean Accuracy: %.3f (%.3f)' % (mean(n_scores), std(n_scores)))
```

Mean Accuracy: 0.931 (0.029)

## After applying Logistic regression Model the Accuracy

```
model3 = LogisticRegression(multi_class='multinomial', solver='lbfgs' , C=100, penalty ="l2",max_iter=10000 ,)
n_scores3 = cross_val_score(model3, X, y, scoring='accuracy', cv=cv, n_jobs=-1)
print('Mean Accuracy: %.3f (%.3f)' % (mean(n_scores3), std(n_scores3)))
#here we train the model and accuracy for this model is 96.7 percent
```

Mean Accuracy: 0.967 (0.028)

# Conclusion

- Successfully predicted whether the user's input is that of a normal or brain cancer patient in this study.
- At the end of our project results will be found with precision.
- Using the required algorithms we will be able to obtain our desired results.
- It will be easier to diagnose after determining whether the patient had a brain cancer gene inside or not.

# BIBLIOGRAPHY

- [1] S.U. Amin, K. Agarwal, R.Beg: Genetic neural network-based data mining in prediction of heart disease using risk factors. In: 2013 IEEE Conference on Information & Communication Technologies (ICT), pp. 1227-1231. (2013)
- [2] A. Ben-Cohen, E. Klang, S. P. Raskin et al., “Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection,” *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.
- [3] F. J. Mateen, M. Carone, N. Alam, P. K. Streatfield, & R. E. Black, (2012). A population- based case-control study of 1250 stroke deaths in rural Bangladesh. *European Journal of Neurology*, 19(7), 999-1006.
- [4] R. Prabhu, Understanding of Convolutional Neural Network (CNN) - Deep Learning., 2019, 21 November [Online]. Available: <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- [5] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, “A deep learning-based framework for automatic brain tumors classification using transfer learning,” *Circuits, Syst., Signal Process.*
- [6] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, “Multi-grade brain tumor classification using deep CNN with extensive data augmentation,” *J. Comput. Sci.*, vol. 30, pp. 174–182, Jan. 2019.

# **Thank You!**