**A Data-Driven Approach to Breast Cancer Prediction: Integrating Machine Learning for Clinical Decision Support.**

First Author1[0000-1111-2222-3333] and Second Author2[1111-2222-3333-4444]

1 Princeton University, Princeton NJ 08544, USA

2 Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany  
lncs@springer.com

**Abstract.** The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words.

**Keywords:** First Keyword, Second Keyword, Third Keyword.

1. **First Section**
   1. **Literature Review**

Shen et al.(2019) developed a CNN model to detect Breast Cancer from biopsies and microscopic images. Here is the efficacy of Shen et al.'s approach was rigorously evaluated on two widely recognized public datasets: the Digital Database for Screening Mammography (DDSM) / CBIS-DDSM and INbreast.For the INbreast dataset, the best single model achieved an impressive per-image AUC of 0.95, with four-model averaging further improving it to 0.98 (sensitivity: 86.7%, specificity: 96.1%) and the DDSM dataset, the best single model achieved a per-image Area Under the Curve (AUC) score of 0.88, which improved to 0.91 with three-model averaging. **[1]**

Siham et al.(2020) focus on how to preprocess data to deal with imbalanced data that have missing values using resampling techniques to enhance the classification accuracy of detecting breast cancer. The three classifiers (NB, SMO, J48) were tested over original data. The accuracy was respectively 71.67%, 69.58%, 75.52%. After applying a discretization filter and removing the records with missing values, results improved. After that, a resample filter was applied for 7 times, then the accuracy was 98.20%(J48), 76.61%(NB), 95.32%(SMO). **[2]**

The study presents to detect Breast cancer that machine learning technique is good enough on linear data. But, when the data form is imaged the machine learning technique fails. For the classification of the breast cancer images data Kumar (2021) claimed that a deep learning based technique CNN gives better results as compared to machine learning techniques. **[3]**

Kumar al(2022) trains machine learning models with some python libraries such as numpy, pandas, matplotlib. They train K-nearest neighbor(KNN), Support vector machine(SVM), Decision tree classifier with that library. They use the Breast malignant growth dataset which was recovered utilizing the UCI archive. The accuracy of those models is 95%(KNN), 96%(SVM) and 93%(DTC). **[4]**

The study introduces the internal functionality of machine learning algorithms. Abien (2018) uses Wisconsin Diagnostic Dataset to train models. GRU-SVM, Linear Regression, MLP, Nearest Neighbor Softmax Regression, SVM algorithms used to train data. The activation function was used for GRU-SVM as like "Sigmoid" or "ReLU". For Nearest Neighbor used Norm function L1 and L2 both. MLP used ReLU as an activation function. Then, trained the models with 128 batch size and 3000 epochs (Nearest Neighbor's is just 1 epoch) . The accuracy of those models was above 90%. **[5]**

### **Sample Heading (Third Level).** Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

#### *Sample Heading (Forth Level).* The contribution should contain no more than four levels of headings. The following Table 1 gives a summary of all heading levels.

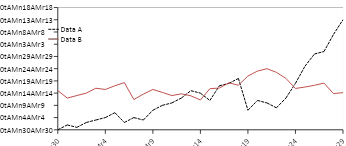
**Table 1.** Table captions should be placed above the tables.

| Heading level | Example | Font size and style |
| --- | --- | --- |
| Title (centered) | **Lecture Notes** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Run-in Heading in Bold.** Text follows | 10 point, bold |
| 4th-level heading | *Lowest Level Heading.* Text follows | 10 point, italic |

Displayed equations are centered and set on a separate line.

*x* + *y* = *z* (1)

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).



**Fig. 1.** A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

**References**

1. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, *9*(1), 12495.
2. Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In *Data Mining and Big Data: 5th International Conference, DMBD 2020, Belgrade, Serbia, July 14–20, 2020, Proceedings 5* (pp. 108-117). Springer Singapore.
3. Priyanka, K. S. (2021). A review paper on breast cancer detection using deep learning. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012071). IOP Publishing.
4. Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 114-118). IEEE.
5. Agarap, A. F. M. (2018, February). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing* (pp. 5-9).