ADScAI
Applied Data Science & Artificial Intelligence Symposium

Deep Ensemble and Uncertainty Estimation for More Accurate Protein Localization in DeepLoc2.0

1st Madusha E.P.C.
Dept. of Computer Science & Engineering
University of Moratuwa
Moratuwa, Sri Lanka
pasindumadusha.20@cse.mrt.ac.lk

2nd Jayathunga A.S.
Dept. of Computer Science & Engineering
University of Moratuwa
Moratuwa, Sri Lanka
anushna.20@cse.mrt.ac.lk

3rd Perera D.N.M.
Dept. of Computer Science & Engineering
University of Moratuwa
Moratuwa, Sri Lanka
navinda.20@cse.mrt.ac.lk

***Keywords*—DeepLoc2.0, Ensemble Learning, Uncertainty Estimation, Protein Sequence Analysis, Bioinformatics**

I. Introduction

Accurately predicting protein subcellular locations is critical in bioinformatics, as it plays a key role in understanding protein functions and their implications in various biological processes. Proteins operate in specific cellular compartments, and their localization is essential for proper function. However, mislocalization can disrupt cellular activities and contribute to the onset of diseases such as metabolic disorders, cardiovascular diseases, neurodegenerative conditions, and cancer. [1] Existing computational models, including deep learning-based approaches like DeepLoc 2.0, have significantly improved localization prediction by leveraging advanced feature extraction techniques. However, a major challenge remains the lack of uncertainty estimation in these models. Given the complexity of protein sequences and the potential for multi-compartment localization, there is an inherent uncertainty in predictions that current models fail to quantify. To address this limitation, we propose a method utilizing deep ensemble learning combined with entropy calculation to mitigate uncertainty in protein localization predictions. By leveraging multiple models and aggregating their predictions, we can achieve higher overall confidence in localization assignments, ensuring more robust and reliable outcomes for protein sequence classification. [2], [3]

II. Literature Review

A. Protein Localization

Protein function is closely tied to its subcellular localization, making accurate localization prediction vital for understanding cellular processes. While traditional methods like homology-based prediction and physicochemical analysis offer insights, they rely heavily on annotated data. In contrast, deep learning models extract features directly from sequences, improving accuracy across diverse localization classes. [2]

B. DeepLoc 2.0

DeepLoc 2.0 represents a significant advancement in deep learning-based protein localization prediction. Unlike its predecessor, DeepLoc 1.0, which relied on convolutional and recurrent neural networks, DeepLoc 2.0 leverages transformer-based architectures to generate more refined vector representations of protein sequences. These transformer models, trained on large-scale protein datasets, capture intricate sequence relationships and structural properties, enhancing prediction accuracy. Additionally, DeepLoc 2.0 incorporates attention mechanisms to identify critical sequence regions associated with sorting signals, further refining the classification process. Its dual-classification framework enables both subcellular localization prediction and sorting signal identification, making it one of the most comprehensive tools for protein localization. [3]

C. Deep Ensembles and Entropy-Based Uncertainty Estimation

Uncertainty estimation in DeepLoc 2.0 predictions is crucial for supporting medical decision-making and guiding further biological research. Reliable confidence in the model's outputs allows researchers to act on its localization predictions with greater assurance. One effective approach to estimate this uncertainty is the deep ensemble technique, which enhances the robustness and trustworthiness of predictions by aggregating outputs from multiple independently trained models. [4], [5]

Entropy-based uncertainty estimation provides a practical way to assess model confidence in DeepLoc 2.0, which is crucial for reliable biomedical applications. Although neural networks are often accurate, they can produce overconfident predictions. Averaging ensemble outputs and computing entropy allows identification of uncertain predictions. By fine-tuning a threshold on entropy values, predictions can be categorized into low- and high-certainty groups. This selective filtering improves interpretability and supports more reliable decision-making in protein localization tasks. [6]

https://doi.org/10.31705/ADScAI.2025.31

III. Materials and Methods

A. Materials

This study leverages two curated protein localization datasets for training and evaluation. The SwissProt dataset (28,303 sequences) from UniProt was used for five-fold cross-validation. Each entry contains a protein sequence (e.g., MNAADRMGARVALLLLVLGSPQSG...) labeled with one of ten true subcellular locations: Cytoplasm, Nucleus, Extracellular, Cell membrane, Mitochondrion, Plastid, Endoplasmic reticulum, Lysosome/Vacuole, Golgi apparatus, or Peroxisome.

To assess the model's generalizability, an entirely independent dataset from the Human Protein Atlas (HPA) was used. This dataset comprises 1,717 sequences that the model has not encountered during training. The model outputs predictions across the same ten-class structure it was trained on. This setup enables a realistic evaluation of DeepLoc 2.0's predictive performance and uncertainty estimation in real-world scenarios, where ground truth annotations may not be available.

B. Methods

The study employed a Deep Ensemble of five DeepLoc 2.0 models, each initialized differently and trained using five-fold cross-validation on the SwissProt dataset. Protein sequences were preprocessed by clipping long inputs and integrating localization annotations, then embedded using the ESM-1b transformer. The model uses an attention mechanism to aggregate token representations, followed by two classifier heads: one for ten-class subcellular localization and another for sorting signal prediction. A weighted focal loss and MCC-based thresholding were used for optimization. To evaluate the model with independent data, the trained ensembles(five) were applied to the unseen Human Protein Atlas (HPA) dataset.

From the five trained models, ensemble predictions were generated for each sample in the HPA dataset, yielding five probability distributions per protein. These were averaged to obtain final class-wise predictions, and predictive entropy was computed by summing the entropies across all ten localization classes for each protein. To identify a confidence-based decision threshold, entropy-based calibration was performed using a held-out subset of SwissProt data, specifically, samples excluded from training in one fold of the cross-validation split. Entropies for these sequences were calculated using the same ensemble procedure. A grid search was then conducted over the entropy range to determine upper and lower uncertainty thresholds by maximizing the Micro-F1 score while maintaining a balanced partitioning of samples into low- and high-uncertainty categories. This trade-off ensured that high-confidence predictions were both reliable and sufficiently represented. Applying the optimized threshold to the HPA predictions resulted in improved performance compared to both individual models and the unfiltered ensemble average. This entropy-guided filtering approach enables researchers to selectively trust DeepLoc 2.0 predictions based on uncertainty levels, supporting more reliable interpretation in downstream biological applications.

TABLE I
RESULT COMPARISON ON HPA DATASET FOR THE ESM-1B MODEL

Metric	Original	Ensemble	Low Uncertainty
Subset Accuracy	0.34	0.34	0.72
Jaccard Index	0.48	0.48	0.82
Micro-F1	0.57	0.57	0.83
Macro-F1	0.44	0.44	0.51
MCC (Cytoplasm)	0.29	0.30	0.31
MCC (Nucleus)	0.41	0.39	0.79
MCC (Cell Membrane)	0.34	0.34	0.50
MCC (Mitochondrion)	0.60	0.60	0.89
MCC (ER)	0.20	0.18	-0.01
MCC (Golgi Apparatus)	0.17	0.23	0.31

IV. Results and Discussion

Table I presents a comparative evaluation of DeepLoc 2.0 using the original model, the deep ensemble, and the entropy-filtered low uncertainty predictions on the HPA dataset.

Despite the observed improvements, this method has certain limitations. The HPA dataset contains a relatively small number of samples, which restricts the ability to achieve a balanced class distribution, especially after filtering based on uncertainty. A significant portion of the data belongs to the Nucleus and Cytoplasm classes, resulting in class imbalance and potential bias in evaluation metrics. Consequently, predictions for less-represented classes are limited, which may affect the reliability of performance estimates in those categories. This imbalance may also reflect underlying limitations in the model's ability to generalize to underrepresented subcellular localizations.

V. Conclusion

By combining deep ensemble learning with entropy-based uncertainty estimation, this study presents a reliable framework for protein subcellular localization, offering biologically meaningful predictions that can support more confident decisions in functional genomics and biomedical research.

References

[1] M.-C. Hung and W. Link, "Protein localization in disease and therapy," *J. Cell Sci.*, vol. 124, pp. 3381–3392, 2011.

[2] J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, and O. Winther, "Deeploc: prediction of protein subcellular localization using deep learning," *Bioinformatics*, vol. 33, pp. 3387–3395, 2017.

[3] V. Thumuluri, J. J. Almagro Armenteros, A. R. Johansen, H. Nielsen, and O. Winther, "Deeploc 2.0: multi-label subcellular localization prediction using protein language models," *Nucleic Acids Res.*, vol. 50, no. W1, pp. W228–W234, 2022.

[4] L. H. Mervin, S. Johansson, E. Semenova, K. A. Giblin, and O. Engkvist, "Uncertainty quantification in drug design," *Drug Discov. Today*, vol. 26, no. 1, pp. 49–59, 2020.

[5] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," *Adv. Neural Inf. Process. Syst.*, 2019.

[6] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 6405–6416.