IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

SNNO: 2347–3657

Volume 9,Issue 4 Nov ,2021

Z-Coordinate Prediction of Residues in _-Helical Transmembrane Proteins Using Deep Learning (TM-ZC)

NILACHAKRA DASH

ABSTRACT

Z-coordinate, defined as the residue's distance from the center of the biological membrane, is a crucial structural property of -helical transmembrane proteins (-TMPs). Neither experimentally solved nor computationally anticipated -TMP structures can z-coordinate prediction allows us to partially describe -TMP structures based on their sequences, which helps with function annotation and drug target finding, and so meets the needs of the relevant study fields. To enhance prediction accuracy and provide a useful tool, we suggested a deep learning-based predictor (TM-ZC) for the z-coordinate of residues in -TMPs. TM-ZC trained a convolutional neural network (CNN) regression model using the one-hot code and the PSSM as input features. The experimental findings showed that TM-ZC was an effective predictor that is both easy to use and quick to run, with respectable results: an average error of 1.958, a percent of prediction error within 3 of 77.461%, and a correlation coefcient (CC) of 0.922. We went on to explore how the TM-ZC predicted z-coordinate may be helpful, and we discovered that it has a high degree of consistency with topological structure and improves the prediction of surface accessibility.

INDEX TERMS

convolutional neural network (CNN), regression, Z-coordinate of residues, -helical transmembrane protein.

INTRODUCTION

The majority of transmembrane proteins (TMPs) consist of -helical structures (TMPs). Using data from UniProt [1], we can see that More than ninety-eight percent of all TMPs are TMPs. Signal transduction [2, 3], nutrition or drug reception [3, 4], immunological response [4, 5], and enzyme activation [5, 6] are just a few of the many functions that -TMPs play in fundamental physiology and pathology. Diseases as diverse as autism [6], epilepsy [7], and cancer [8][11] may have their origins in -TMP malfunction. Therefore, more than 50% of all TMPs aim towards

ASSISTANT PROFESSOR, Mtech,Ph.D
Department of CSE
Gandhi Institute for Technology,Bhubaneswar.

The effectiveness of drug development depends on a thorough understanding of the structure of existing medications [12], [13]. Despite their vital biological roles, however, determination.

Due to ongoing technological challenges, only around 5% of -TMPs have had their high-resolution structures identified.

As a result, various structural descriptors extracted from original sequences are being used to boost TMP-related research efforts in the present. Topology structure, surface accessibility, and z-coordinate are all examples of low-resolution structural descriptors that may complement high-resolution structural data for learning about -TMPs. Many novel approaches to illumination have been developed in recent years and have made significant progress. This includes VOLUME 8 and the like.

With the goal of improving upon previous techniques for predicting the topological structure of -TMPs [14, 15], S. H. Feng et al. rst created a multiscale deep learning protocol (MemBrain 3.0) that has two distinct layers of neural networks modular components; transmembrane helix prediction and orientation prediction [16]. In a similar vein, various approaches have been developed to estimate the surface accessibility of -TMPs, with some of them achieving very good performance [17], [18]. For instance, in our prior work [19] we introduced a deep learning-based predictor (TMP-SSurface) that used one-hot codes and PSSM as input characteristics to create a hybrid of the Inception and the CapsuleNet.

The distance from a residue in -TMP to the membrane's geometric center is defined as the residue's Z-coordinate [20].

Z-coordinate, like topological structure, reflects the connection between the residue and the membrane, but it does so via continuous numerical measurement. Since ligand-binding and protein-protein binding sites are usually in very particular places on transmembrane, water-soluble, or junction regions, the z-coordinate is significantly connected with them. Topology prediction [21], structural classication [22], burial status prediction [23], and many more fields may all benefit from the anticipated z-coordinate [24], [25]. Computational approaches that accurately anticipate the z-coordinate of residues in -TMPs are not only a necessary step towards structure identification, but also have the potential property of aiding in function annotation, drug target research, and other related issues. [21], [26], [27].

But research into the z-coordinate has lagged behind that of topological structure and surface accessibility.

Only one z-coordinate predictor, ZPRED [20], has been published in the last decade; it combines an Artificial Neural Network (ANN) with a Hidden Markov Model (HMM) and takes sequential information as inputs. If you're looking for the original work on z-coordinate prediction, ZPRED is it, however its website has since been taken down. In order to further -TMP studies, a z-coordinate predictor that is both accurate and fast is required.

The number of buildings that contain -TMPs has grown by a factor of more than 10 in the previous 15 years. The study might benefit from additional data, and the deep learning approach offers a fresh way to build a more effective predictor that is easier to implement and runs more quickly without sacrificing accuracy.

In this paper, we put forward a deep learning-based predictor (TM-ZC) for the z-coordinate of residues in -TMPs. TMZC trained a convolutional neural network (CNN) regression model using the one-hot code and the PSSM as input features.

Experiments showed that TM-ZC performed well. The CC was 0.917, the mean error was 1.865, and the percentage of incorrect predictions within 3 standard deviations was 76.703. Additionally, we examined the potential of the TM-ZC predicted z-coordinate in solving the issues of surface accessibility prediction and topological structure prediction. By setting a lower limit on the z-coordinate predicted by TM-ZC, we sought to

separate the transmembrane residues from the non-transmembrane residues.

The z-coordinate predicted by TM-ZC is well correlated with the topological structure, as proved by the experiments. Predicting surface accessibility requires.

Additionally to our earlier study, we have included the TM-ZC projected z-coordinate in order to foretell the surface accessibility of -TMPs. Predicted z-coordinates from TM-ZC improved prediction performance experimentally. http://icdtools.nenu.edu.cn/TM-ZC provides a dependable webserver that anybody is welcome to use.

II. MATERIALS AND METHODS

BENCHMARK DATASETS

In 2005, a dataset of 101 non-homologous chains from 46 complexes was produced for ZPRED's usage.

A more comprehensive benchmark dataset is what we think important urged. Most researchers rely on the information found in the Protein Data Bank of Transmembrane Proteins (PDBTM) [28]. To generate it, we ran the TMDET algorithm [29] over every entry in the PDB. PDBTM data shows that in the last 15 years, the number of -TMPs has grown by a factor of more than 10. From PDBTM, we retrieved a total of 3820 complexes including 13,209 -TMP sequences (version: 2019-05-10). Getting rid of sequences that use atypical amino acid residues Short sequences of less than 30 residues were omitted since they were always interpreted as peptides. To mitigate homology bias's deleterious effects [30], we ran CD-HIT with a 0.3 sequence identity cut-off to cluster the remaining proteins, and then we extracted the longest sequences from each cluster. After initial processing, 851 -TMPs containing a total of 223,310 residues were recovered. To verify TM-robustness, ZC's a random sample of 50 sequences was chosen and used in an independent test (ZC-test50). Only 50 sequences were utilized as the validation dataset (ZC-valid50), while the other 751 sequences were used to construct the training and tuning datasets for the prediction model (ZC-train751). In ten-fold cross validation, we repeated the procedure of choosing the validation dataset 10 times. When training the models, the results shown below are an average of the results from 10 rounds of cross-validation on each of the sub-models. All datasets utilized for this study may be found in the Additional Files

CALCULATION OF Z-COORDINATE

Z-COORDINATE CALCULATION

The residues' PDB coordinates must be rotated and shifted from their original positions to account for relative membrane and protein-specific locations. The value of the z-coordinate as seen may be determined using Formula 1:

$$\left[x_i', y_i', z_i'\right] = [x_i, y_i, z_i]A^{-1} + [b_x, b_y, b_z], \qquad (1)$$

A is the matrix that rotated the protein around the origin, and [xi; yi; zi] are the coordinates of the alpha-carbon atom of the ith residue as they were in the PDB les downloaded from PDBTM antiparallel to the membrane's surface. (bx), (by), (bz) is a transporter that delivered the protein to its membrane destination.

TMDET[29] was the source for both A and bx ; by ; bz. The membrane's nucleus is located at a z-coordinate of 0.

Then, two threshold cutting phases were conducted, mirroring the procedure of ZPRED:

First, we ignored orientation and only looked at the distance separating the residues from the center of the membrane (z0 I which limited the threshold from.1;C1/ to [0;C1). The formula for the absolute value is:V

$$z_{i[0,+\infty)} = \left|z_i'\right|, \qquad (2)$$

Z-coordinate values between 0 and 5 were designated as belonging to a core hydrophobic area, while all values over 25 were designated as non-transmembrane residues. Observed TM-ZC z-coordinate labels were only those that fell within

International Journal of Information Technology & Computer Engineering

SNNO: 2347–3657

Volume 9,Issue 4 Nov ,2021

the range [5; 25]. You may figure it out using the formula:V

$$z_{i[5,25]} = \begin{cases} 5, & \left|z_i^{'}\right| \le 5 \\ \left|z_i^{'}\right|, & 5 < \left|z_i^{'}\right| < 25 \\ 25, & 25 \le \left|z_i^{'}\right|, \end{cases} \quad (3)$$

Encoding of Protein Sequences (C.)

1) CONSERVATION BY EVOLUTION (PSSM)

Particular genetic traits of proteins have been rising in popularity, especially among proteins with similar structures. High-conservation protein fragments have been shown to be directly connected to the proteins' structural or functional requirements [31, 32]. Among the useful descriptors derived from multiple sequence alignment is the position-specific score matrix (PSSM) [33]. PSSMs were determined by running PSI-BLAST [34] on the UniRef50 database (published on October 16, 2019) with an e-value cutoff of 0.001 and 3 iterations. Protein secondary structure motif (PSSM) defined as 20 L matrix

$$PSSM = \begin{bmatrix} P_{1,AA_1} & P_{1,AA_2} & \cdots & P_{1,AA_{20}} \\ P_{2,AA_1} & P_{2,AA_2} & \cdots & P_{2,AA_{20}} \\ \vdots & \vdots & \vdots & \vdots \\ P_{L,AA_1} & P_{L,AA_2} & \cdots & P_{L,AA_{20}} \end{bmatrix}, \quad (4)$$

where Pi;AAj is the PSSM element value, where PSSM is the probability that AAj appears at position I of the given protein in a multiple sequence alignment.

The length of the protein, denoted by L. In the next step, we utilized the logistic function to transform each PSSM value into the interval [0, 1]:

$$P_{i,AA_j}^{'} = \frac{1}{1 + e^{-P_{i,AA_j}}}, \quad (5)$$

2) SINGLE-USE CODE

For each residue in a protein sequence, one-hot coding is employed to express its kind using a sparse encoding scheme. This is the single most simplest approach of describing a protein's sequence, using just the 20 standard amino acids and their relative positions. Deep learning-based protein function predictions have shown this to be a useful characteristic [35, 39]. A 20 L matrixV represents the protein's one-hot coding.

$$one-hot = \begin{bmatrix} O_{R_1,AA_1} & O_{R_1,AA_2} & \cdots & O_{R_1,AA_{20}} \\ O_{R_2,AA_1} & O_{R_2,AA_2} & \cdots & O_{R_2,AA_{20}} \\ \vdots & \vdots & \vdots & \vdots \\ O_{R_L,AA_1} & O_{R_L,AA_2} & \cdots & O_{R_L,AA_{20}} \end{bmatrix}, \quad (6)$$

where *ORi;AAj* represents the element's value of one-hot code.

Ri is the type of residue on position *i*. *AAj* is the type of 20 standard amino acids. *ORi;AAj* D 1 if *Ri* D *AAj*; *ORi;AAj* D 0 if *Ri* 6D *AAj*. *L* represents the length of the protein. ORi;AAj is the one-hot code for that element.

Ri represents the residue type at position i. There are 20 typical amino acids in AAj. If (Ri = AAj), then (ORi;AAj = 1) ORi:AAj=D0 if Ri:6D The protein's length, denoted by the letter L, is AAj.

D. ORIGINAL MODELING

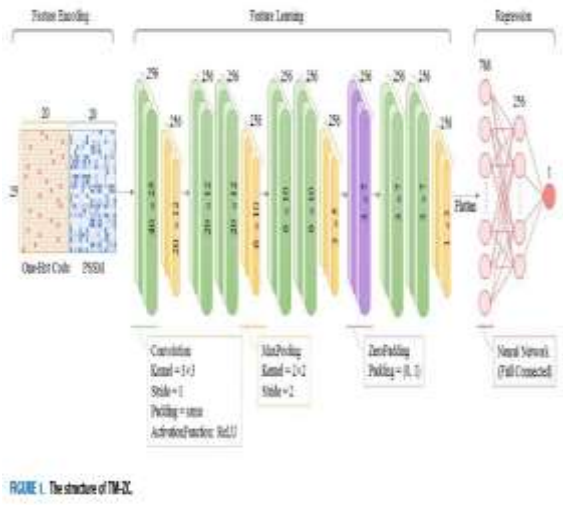
To put it simply, a convolutional neural network (CNN) is a feedforward neural network in which the neurons are able to reflect the information that is sent into them.

information around the area covered by the convolution kernel. In each iteration, the network was trained using the training dataset, and then its performance was evaluated using the validation dataset, which was then utilized to provide feedback for the training process in the following iteration. The need for human-created characteristics is eliminated, and instead, relevant traits may be gleaned straight from raw data. CNN excels in the fields of image/video recognition [42], natural language processing [43], and medical diagnosis [44]. CNN's efficacy has led to its widespread use in bioinformatics, particularly in tasks like super-enhancer prediction [45] and drug-disease association prediction [46]. Our aim in this

study was to simplify the underlying prediction model.

Our solution was to build a simple network structure. Figure 1 depicts the layout of the prediction model. For activation, we used ReLU, and each convolution layer included 256 kernels of size 3, stride 1, and ReLU.

Maximal pooling was used on all 256 kernels across all pooling layers, and the kernel size and stride were also set to 2. Features were extracted from the data using a single convolution layer in this model.



in its raw form, the feature matrix was encoded. After that, a single layer of max pooling was executed. Two convolution layers, a max pooling layer, another pair of convolution layers, and

Each layer of the max pooling procedure was executed in turn. The feature matrix size was 35, which was insufficient to run the subsequent layers. Consequently, we added a zero-padding layer to the features matrix to increase their size from 3 by 7 to 20 by 20. After that, a maximum pooling layer and a pair of convolution layers were executed. Currently, feature extraction is performed using 256 x 1 3 matrices. We lined them up, and then ran a neural network with all of its connections active. Seven hundred and sixty-eight neurons make up the input layer, 256 neurons are in the hidden layer, and a single neuron is found in the output layer.

With this prediction, we took into account the potential presence of an output neuron-related target residue.

EVALUATION OF WORK PRODUCTIVITY

Mean absolute error, Pearson correlation coefficient, and percentage of correct findings (P3) were used to evaluate TM-prediction ZC's performance. The mean absolute error (MAE) is the sum of the differences between the calculated and measured z-coordinates of all residues. The lesser the MAE number, the greater the performance. Its range was [0, 1].

The coefficient of determination (CC) displays the linear relationship between the calculated and measured z-coordinate. The closer the CC was to 1 in the range [1, 1], the better it performed. The optimum prediction ratio, P3, is the threshold taken from ZPRED. Generally speaking, the bigger the ratio was, the better it performed, with a range of [0%, 100%].

$$\mathrm{MAE} = \frac{1}{L}\sum_{i=1}^{L} |y_i - x_i|, \tag{7}$$

$$\mathrm{CC} = \frac{\sum_{i=1}^{L} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{L} (x_i - \bar{x})^2\right]\left[\sum_{i=1}^{L} (y_i - \bar{y})^2\right]}}, \tag{8}$$

40132

$$P_{3\text{Å}} = \frac{N_{|y_i - x_i| < 3\text{Å}}}{L}, \tag{9}$$

Nx and Ny are the respective median values for the observed and anticipated z-coordinates of the ith residue, respectively; L is the total number of residues. Njyi?xij<3Å quantifies the fraction of residues for which the prediction error is below 3.

III. CONCLUSIONS AND RECOMMENDATIONS

PRODUCT FEATURE ANALYSIS

We conducted an ablation research on features to learn more about the impact of various types of

features and their role in the prediction model. There are three distinct models since we used the one-hot code, the PSSM, and both.

TABLE 1 demonstrates that PSSM is superior than one-hot code.

The phenomenon exemplifies the tight connection between protein structure and evolutionary conservation. Although the model's performance suffered when employing just the one-hot code feature, it improved when combined uring traininwith the PSSM feature.

B. IMPACT OF WINDOWSIZE

In this study, we adopted a sliding window method, and we found that the prediction performance of TM-ZC was very sensitive to the value of the window size. We experiment with a range of window sizes, from 15 to 31, with a step size of 2. How well TM-ZC was able to predict on the validation dataset dg

TABLE 1. Performance comparison of the different models on the feature ablation study.

window size	MAE	CC	$P_{3\lambda}$ (%)
One-hot code	8.801	0.675	32.540
PSSM	4.598	0.873	52.292
One-hot code+PSSM	**4.373**	**0.917**	**55.343**

The bolded parts represent the highest value of the corresponding evaluation indicator.

TABLE 2. The prediction performance by using different window sizes.

window size	MAE	CC	$P_{3\lambda}$ (%)
15	5.267	0.883	47.833
17	4.918	0.894	50.073
19	1.827	0.924	78.66400
21	1.726	0.927	79.12000
23	1.755	0.925	78.80800
25	**1.673**	**0.939**	**79.10700**
27	1.741	0.93	79.69300
29	1.838	0.926	79.00700
31	1.882	0.922	80.05700

TABLE 2 shows the results of experimenting with various window widths.

TM-performance ZC's rose steadily as the window size increased, peaking with the value of size of window topped out at 25. Consequently, in all our tests, we used a window size of 25.

C. THE RESULT OF SLASHING THE THRESHOLD

The original residue coordinates recorded in the PDB data were rotated and relocated according to the relative locations of the protein and the membrane, as detailed in "Section II-BCalculation of Z-coordinate." Two threshold-cutting iterations were then carried out. In the rst stage, we reduce the threshold from.1;C1/ to [0;C1] using the absolute value. The second process set a lower bound for the value, between [5 and 25]. After obtaining three sets of training labels with varying cutoffs, three distinct models were developed. TABLE 3 shows how well these models performed on the validation dataset (ZC-valid50).

The model trained with the original labels and a threshold of.1;C1/ clearly underperformed, whereas the other two models showed substantial improvement.

As a potential explanation for this phenomena, there is one that has been proposed:

It is challenging for the prediction model to determine the link between the characteristics and the labels for residues that are symmetrical around the membrane center since they always share the same attributes but have opposing labels. Another observation is that models trained on labels with a threshold of [5; 25] performed the best.

TABLE 3. The Prediction performance of models, which are trained by using training labels with different cutting thresholds.

Cutting Threshold	MAE	CC	$P_{3\lambda}$ (%)
Performance on ZC-valid50 dataset			
$(-\infty, +\infty)$	30.067	0.010	17.080
$[0, +\infty)$	6.077	0.903	45.266
[5,25]	**1.65**	**0.931**	**80.109**

International Journal of Information Technology & Computer Engineering

SNNO: 2347–3657

Volume 9,Issue 4 Nov ,2021

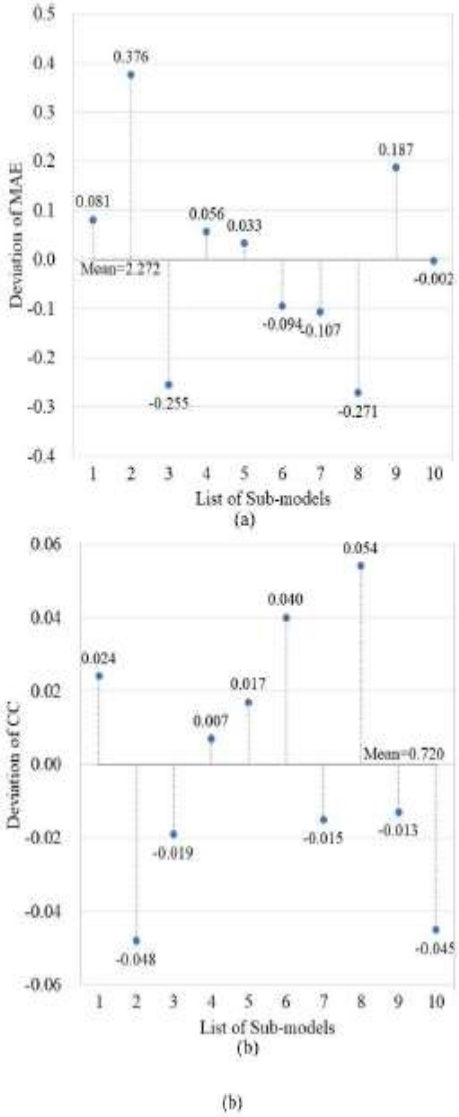


FIGURE 2. The stability of the training process. (a) The performance of MAE on the cross validation. (b) The performance of CC on the cross validation.

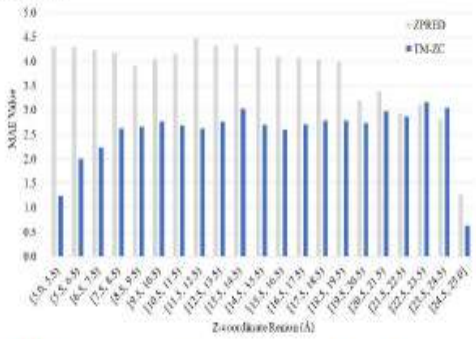


FIGURE 3. Comparison of MAE in different z-coordinate regions.

D. PERFORMANCE OF TM-ZC

In order to investigate the performance of TM-ZC and verified its stability, we performed ten-fold cross validation.

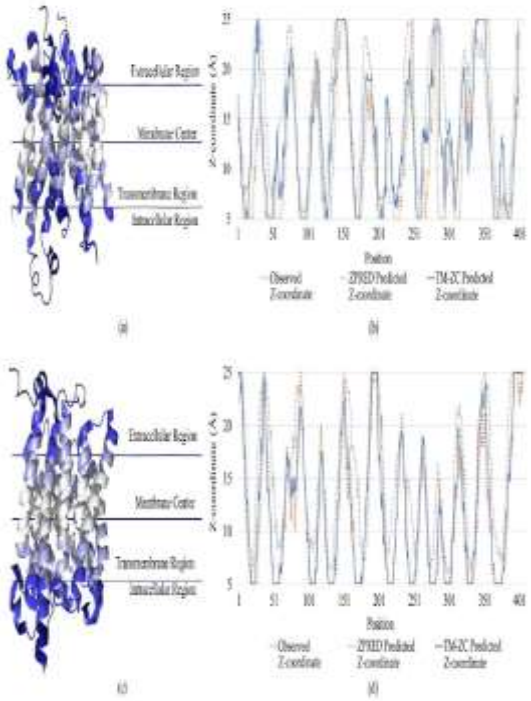


FIGURE 4. Case studies of TM-ZC: took 4E0A_A and 5L25_A as examples. (a) and (c) are 3D visualization of the prediction result of 4E0A_A and 5L25_A, respectively. As the value of predicted z-coordinate increases, the color changes from white to dark blue. (b) Illustrate the curves of the z-coordinate value of 4E0A_A: the observed value (orange dotted curve), the ZPRED predicted value (gray dotted curve), and the TM-ZC predicted value (blue curve). (d) Illustrate the curves of the z-coordinate value of 5L25_A, which is as same as (b).

Figure 2 displays the TM-effectiveness ZC's in terms of (a) MAE value and (b) CC value, respectively. From (a), we can see that the average MAE across 10 models is 2.215, with a standard deviation of

each sub-mean model's absolute error and its mean value are shown on the label. In (b), we see that across 10 models, the average CC is 0.915; the difference between each sub-CC model's and the mean is shown as the label value. Figure 2 shows that TM-performance ZC's was consistent throughout all cross-validation tests.

The sole predictor is ZPRED, thus it's important to compare results with it. To test the TM-efficacy, ZC's we compared it to ZPRED.

Comparison of the mean absolute error (MAE) between ZPRED and TM-ZC in various z-coordinate locations is shown in Fig. 3. It was clear that, in terms of performance, TM-ZC was superior than

ZPRED throughout the board, but notably at the low end of the z-coordinate.

The residue is in the hydrophobic transmembrane region if the z-coordinate is low. It showed that TM-ZC worked well for transmembrane residues.

In the section [19:5; 24:5], TM-benefits ZC's were obscured and it performed worse than ZPRED. This demonstrates the necessity for further refinement of TM-prediction ZC's ability for residues located in the junction region of the membrane surface.

TABLE 4. Comparison of the performance between ZPRED and TM-ZC.

Predictor	MAE	CC	$P_{3\AA}$ (%)
ZPRED	3.293	0.813	59.731
TM-ZC	1.958	0.922	77.461

To see how ZPRED and TM-ZC stack up against one another in terms of overall prediction performance, please refer to TABLE 4. Clearly TM-ZC did better than ZPRED. A decrease in the MAE of More than 28% more results had an error of 3 or less (P3), and the Pearson correlation coefficient (CC) was raised from 0.0 to 0.1, for a total increase of around 43%.

D. CASE STUDIES

To further prove how useful TM-ZC is, we conducted case studies. Examples are taken from ZC-test50, with 6EU6 A and 5L25 A being selected. The protein 6EU6 A was isolated from Escherichia coli and has eleven transmembrane domains. ATP, Dodecyl-Alpha-D-Maltoside, and other ligands bind to it since it is their target. Saccharomyces Cerevisiae 5L25 A is a ten-transmembrane protein.

It's crucial to anion exchange and borate transfer. As shown in Fig. 4, the outcomes of the predictions for both proteins.

In Fig. 4: (a) depicted the z-coordinate that TM-ZC predicted for 6EU6 A, with darker colors indicating greater z-coordinate values. In (b), the z-coordinate curve value, the predicted value using ZPRED, and the predicted value using TM-ZC, all for 6EU6 A. Similar to (a) and (b), (c) and (d) depicted the data of 5L25 A. (b).

From (a) and (c), it's clear that TM-anticipated ZC's z-coordinate is in good agreement with the data.

In addition, (b) and (d) confirm that the TM-ZC projected value curve closely ts the actual value curve, making it superior than ZPRED.

G. THE TOPOLOGICAL STRUCTURE IS CORRELATED WITH THE Z-COORDINATE

The residue topological structure of TMPs is intimately related to the z-coordinate. By setting a lower limit on the z-coordinate predicted by TM-ZC, we sought to separate the transmembrane residues from the non-transmembrane residues.

Experiments conducted on the ZC-test50 dataset showed that a threshold of 14.5 resulted in the maximum accuracy (79.268%). Transmembrane residues were defined as those having a z-coordinate less than or equal to 14.5. Since the biological experiment demonstrated that the mean thickness of the membrane is around 30, this would indicate that the mean thickness of the membrane was 14:5 2 D 29. Because of this, there is a robust association between topological structure and z-coordinate, and the threshold of TMZC predicted z-coordinate is consistent with facts.

H. Improvements in surface accessibility prediction enabled by TM-ZC

The TM-ZC predicted z-coordinate has a high degree of agreement with the topological structure, and it may also improve predictions of the surface accessibility of -TMPs residues. Once upon a time, we put out a predictor (TMPSSurface) for predicting residues' relative accessible surface areas (rASAs) in -TMPs [19]. TMP-SSurface was a deep learning-based regression technique that used as input features one-hot code, terminal ag, and PSSM. To further validate TM-use, ZC's we included the algorithm's projected z-coordinate as an extra feature and were pleased to see that the CC value climbed from 0.581 to 0.604. The results

of the experiment demonstrated the existence of a connection between z-coordinates and rASA.

Final Thoughts

In -TMPs, the z-coordinate of a residue is defined as the residue's distance from the membrane's geometric center. It's a useful structural description that has a strong relationship to the known functional domains of -TMPs. ZPRED is the only existing predictor for this issue, hence it obviously needs to be refined. Predicting the z-coordinate of residues in -TMPs may be difficult, thus we've suggested a deep learning-based predictor (TM-ZC). Using one-hot code and PSSM as input characteristics, TM-ZC is a straightforward CNN-based predictor. With an MAE of 1.958 and a CC of 0.9, TM-ZC performed well.

A prediction error of 77.461% was found to be within a 3-standard-deviation margin of error for a 0.922 value. In experiments, we observed the impact of two feature types and found that PSSM features were more effective.

We also checked the accuracy of the predicted z-coordinate from TM-ZC and its applicability to the issues of surface accessibility prediction.

The results of these experiments show that TM-ZC may be a useful tool for addressing these issues. We have faith that TM-ZC will be useful in future studies of other types of transmembrane proteins.

REFERENCES

[1] T. UniProt Consortium, ``UniProt: A worldwide hub of protein knowledge,'' *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506_D515, Jan. 2019, doi: 10.1093/nar/gky1049.

[2] L. He, E. B. Cohen, A. P. B. Edwards, J. Xavier-Ferrucio, K. Bugge, R. S. Federman, D. Absher, R. M. Myers, B. B. Kragelund, D. S. Krause, and D. DiMaio, ``Transmembrane protein aptamer induces cooperative signaling by the EPO receptor and the cytokine receptor _-common subunit,'' *iScience*, vol. 17, pp. 167_181, Jul. 2019, doi: 10.1016/j.isci.2019. 06.027.

[3] M. L. Colgrave, K. Byrne, S. V. Pillai, B. Dong, A. Leonforte, J. Caine, L. Kowalczyk, J. A. Scoble, J. R. Petrie, S. Singh, and X.-R. Zhou, ``Quantitation of seven transmembrane proteins from the DHA biosynthesis pathway in genetically engineered canola by targeted mass spectrometry,'' *Food Chem. Toxicol.*, vol. 126, pp. 313_321, Apr. 2019, doi: 10.1016/j.fct.2019.02.035.

[4] L. Chen, Y. Zhang, S. Zhang, Y. Chen, X. Shu, J. Lai, H. Cao, Y. Lian, Z. Stamataki, and Y. Huang, ``A novel T-cell epitope in the transmembrane region of the hepatitis B virus envelope protein responds upon dendritic cell expansion,'' *Arch. Virology*, vol. 164, no. 2, pp. 483_495, Feb. 2019, doi: 10.1007/s00705-018-4095-0.

[5] X. Duan, X. Liao, S. Li, Y. Li, M. Xu, Y. Wang, H. Ye, H. Zhao, C. Yang, X. Zhu, and L. Chen, ``Transmembrane protein 2 inhibits zika virus replication through activation of the janus kinase/signal transducers and activators of transcription signaling pathway,'' *Future Virol.*, vol. 14, no. 1,pp. 9_19, Jan. 2019, doi: 10.2217/fvl-2018-0115.

[6] S. K. Ra_, A. Fernández-Jaén, S. Álvarez, O. W. Nadeau, and M. G. Butler, ``High functioning autism with missense mutations in synaptotagmin-like protein 4 (SYTL4) and transmembrane protein 187 (TMEM187) genes: SYTL4-protein modeling, protein-protein interaction, expression pro_ling and MicroRNA studies,'' *Int. J. Mol. Sci.*, vol. 20, no. 13, p. 3358, Jul. 2019, doi: 10.3390/ijms20133358.

[7] Y. Tanabe, T. Taira, A. Shimotake, T. Inoue, T. Awaya, T. Kato, A. Kuzuya, A. Ikeda, and R. Takahashi, ``An adult female with proline-rich transmembrane protein 2 related paroxysmal disorders manifesting paroxysmal kinesigenic choreoathetosis and epileptic seizures,'' *Rinsho Shinkeigaku*, vol. 59, no. 3, pp. 144_148, Mar. 2019, doi: 10.5692/clinicalneurol.cn-001228.

[8] Y. Moon, W. Lim, and B. Jeong, ``Transmembrane protein 64 modulates prostate tumor progression by regulating Wnt3a secretion,'' *Oncol. Lett.*, vol. 18, no. 1, pp. 283_290, Jul. 2019, doi: 10.3892/ol.2019.10324.

[9] D. Tao, J. Liang, Y. Pan, Y. Zhou, Y. Feng, L. Zhang, J. Xu, H. Wang, P. He, J. Yao, Y. Zhao, Q. Ning, W. Wang, W. Jiang, J. Zheng, and X. Wu, ``*In vitro* and *in vivo* study on the effect of lysosome-associated protein transmembrane 4 beta on the progression of breast cancer,'' *J. Breast Can- cer*, vol. 22, no. 3, pp. 375_386, Sep. 2019, doi: 10.4048/jbc.2019.22.e43.

[10] J. Yan, Y. Jiang, J. Lu, J. Wu, and M. Zhang, ``Inhibiting of proliferation, migration, and invasion in lung cancer induced by silencing interferoninduced transmembrane protein 1 (IFITM1),'' *BioMed Res. Int.*, vol. 2019, pp. 1_9, May 2019, doi: 10.1155/2019/9085435.

[11] K. Qu, F. Gao, F. Guo, and Q. Zou, ``Taxonomy dimension reduction for colorectal cancer prediction,'' *Comput. Biol. Chem.*, vol. 83, Dec. 2019, Art. no. 107160, doi: 10.1016/j.compbiolchem.2019.107160.

[12] T. Langó, G. Róna, É. Hunyadi-Gulyás, L. Turiák, J. Varga, L. Dobson, G. Várady, L. Drahos, B. G. Vértessy, K. F. Medzihradszky, G. Szakács, and G. E. Tusnády, ``Identi_cation of extracellular segments by mass spectrometry improves topology prediction of transmembrane proteins,'' *Sci. Rep.*, vol. 7, no. 1, Feb. 2017, Art. no. 42610, doi: 10.1038/srep42610.

[13] L. Yu, X. Sun, S. W. Tian, X. Y. Shi, and Y. L. Yan, ``Drug and nondrug classi_cation based on deep learning with various feature selection strategies,'' *Current Bioinf.*, vol. 13, no. 3, pp. 253_259, 2018, doi: 10.2174/1574893612666170125124538.

[14] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, ``Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes,'' *J. Mol. Biol.*, vol. 305, no. 3, pp. 567_580, Jan. 2001, doi: 10.1006/jmbi.2000.4315.

[15] H. Wu, K. Wang, L. Lu, Y. Xue, Q. Lyu, and M. Jiang, ``Deep conditional random _eld approach to transmembrane topology prediction and application to GPCR three-dimensional structure modeling,'' *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 5, pp. 1106_1114, Sep. 2017, doi: 10.1109/TCBB.2016.2602872.

[16] S. H. Feng, W. X. Zhang, J. Yang, Y. Yang, and H. B. Shen, ``Topology prediction improvement of alpha-helical transmembrane proteins through helix-tail modeling and multiscale deep learning fusion,'' *J. Mol. Biol.*, vol. 432, no. 4, pp. 1279_1296, Dec. 2019, doi: 10.1016/j.jmb.2019.12.007.

[17] K. Illergard, S. Callegari, and A. Elofsson, ``MPRAP: An accessibility predictor for a-helical transmem-brane proteins that performs well inside and outside the membrane,'' *BMC Bioinf.*, vol. 11, no. 1, p. 333, Jun. 2010, doi: 10.1186/1471-2105-11-333.

[18] X. Yin, J. Yang, F. Xiao, Y. Yang, and H.-B. Shen, ``MemBrain: An easyto- use online webserver for transmembrane protein structure prediction,'' *Nano-Micro Lett.*, vol. 10, no. 1, 2018, Art. no. 2, doi: 10.1007/s40820- 017-0156-2.

[19] C. Lu, Z. Liu, B. Kan, Y. Gong, Z. Ma, and H. Wang, ``TMP-SSurface: A deep learning-based predictor for surface accessibility of transmembrane protein residues,'' *Crystals*, vol. 9, no. 12, p. 640, Dec. 2019,doi: 10.3390/cryst9120640.

[20] E. Granseth, H. Viklund, and A. Elofsson, ``ZPRED: Predicting the distance to the membrane center for residues in -helical membrane proteins,'' *Bioinformatics*, vol. 22, no. 14, pp. e191_e196, Jul. 2006, doi: 10.1093/bioinformatics/btl206.

[21] A. Bernsel, H. Viklund, A. Hennerdal, and A. Elofsson, ``TOPCONS: Consensus prediction of membrane protein topology,'' *Nucleic Acids Res.*, vol. 37, pp. W465_W468, May 2009, doi: 10.1093/nar/gkp363.

[22] H. Viklund, E. Granseth, and A. Elofsson, ``Structural classi_cation and prediction of reentrant regions in _-helical transmembrane proteins: Application to complete genomes,'' *J. Mol. Biol.*, vol. 361, no. 3, pp. 591_603, Aug. 2006, doi: 10.1016/j.jmb.2006.06.037.

[23] Y. Park, S. Hayat, and V. Helms, ``Prediction of the burial status of transmembrane residues of helical membrane proteins,'' *BMC Bioinf.*, vol. 8, no. 1, p. 302, Aug. 2007, doi: 10.1186/1471-2105-8-302.

[24] C. Papaloukas, E. Granseth, H. Viklund, and A. Elofsson, ``Estimating the length of transmembrane helices using Z-coordinate predictions,'' *Protein Sci.*, vol. 17, no. 2, pp. 271_278, Feb. 2008, doi: 10.1110/ps.073036108.

[25] B. Wallner, ``ProQM-resample: Improved model quality assessment for membrane proteins by limited conformational sampling,'' *Bioinformatics*, vol. 30, no. 15, pp. 2221_2223, Apr. 2014, doi: 10.1093/bioinformatics/ btu187.

[26] T. Nugent and D. T. Jones, ``Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis,'' *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 24, pp. E1540_E1547, May 2012, doi: 10.1073/pnas.1120036109.

[27] A. Rose, S. Lorenzen, A. Goede, B. Gruening, and P. W. Hildebrand, ``RHYTHM_a server to predict the orientation of transmembrane helices in channels and membrane-coils,'' *Nucleic Acids Res.*, vol. 37, pp. W575_W580, May 2009, doi: 10.1093/nar/gkp418.

[28] D. Kozma, I. Simon, and G. E. Tusnády, ``PDBTM: Protein data bank of transmembrane proteins after 8 years,'' *Nucleic Acids Res.*, vol. 41, no. D1, pp. D524_D529, Jan. 2013, doi: 10.1093/nar/gks1169.

[29] G. E. Tusnady, Z. Dosztanyi, and I. Simon, ``TMDET: Web server for detecting transmembrane regions of proteins by using their 3D coordinates,'' *Bioinformatics*, vol. 21, no. 7, pp. 1276_1277, Apr. 2005, doi: 10.1093/bioinformatics/bti121.