Deep learning in prediction of intrinsic disorder in proteins

Bi Zhao, Lukasz Kurgan *

Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA

ABSTRACT

Intrinsic disorder prediction is an active area that has developed over 100 predictors. We identify and investigate a recent trend towards the development of deep neural network (DNN)-based methods. The first DNN-based method was released in 2013 and since 2019 deep learners account for majority of the new disorder predictors. We find that the 13 currently available DNN-based predictors are diverse in their topologies, sizes of their networks and the inputs that they utilize. We empirically show that the deep learners are statistically more accurate than other types of disorder predictors using the blind test dataset from the recent community assessment of intrinsic disorder predictions (CAID). We also identify several well-rounded DNN-based predictors that are accurate, fast and/or conveniently available. The popularity, favorable predictive performance and architectural flexibility suggest that deep networks are likely to fuel the development of future disordered predictors. Novel hybrid designs of deep networks could be used to adequately accommodate for diversity of types and flavors of intrinsic disorder. We also discuss scarcity of the DNN-based methods for the prediction of disordered binding regions and the need to develop more accurate methods for this prediction.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Intrinsic disorder in proteins is defined by lack of stable tertiary structure under physiological conditions [1–4]. Intrinsically disordered proteins (IDPs) include one or more intrinsically disordered regions (IDRs) in their sequences. Recent bioinformatics investigations conclude that IDPs are highly abundant in eukaryotic organisms [5–7] and enriched in multiple cellular compartments [8,9]. Numerous studies of IDPs reveal that they are crucial for a wide spectrum of cellular functions that include signaling, molecular recognition and assembly, cell cycle regulation, transcription, translation and phase separation [10–19]. Moreover, given their functional importance and prevalence in the human diseasome [12,20–22], they serve as promising and currently underutilized leads for rational drug design efforts [23–27].

Experimentally characterized IDPs and IDRs can be collected from several databases, such as DisProt [28], PDB [29], IDEAL
PrDOS [31], DIBS [31], and MFbI [32]. However, these resources cover only a small fraction of IDPs, with the largest DisProt and PDB databases currently including about 2 thousand and 25 thousand IDPs, respectively [28,33]. Compared to over 225 million protein sequences that are available in the newest 2021_04 release of UniProt [34], we have a long way to go to comprehensively identify and annotate IDPs and IDRs. Computational methods that accurately predict intrinsic disorder can be used to facilitate efforts to close this huge and growing knowledge gap. Computational predictors already made large impact on the intrinsic disorder field, by powering a rapid acceleration in the research on IDPs and IDRs [35]. They are also used across many areas including rational drug design [23–26], structural genomics [36–38], and medicine [39,40].

Development of computational predictors of disorder is a long-standing research problem. A recent survey has identified 103 disorder predictors that were developed over the last four decades [41]. Current surveys point to the long history of the disorder prediction area, providing invaluable insights concerning architectures of these methods, their availability, trends in their development efforts and approaches to comparatively evaluate their predictive performance [40–48]. Moreover, users and developers benefit from empirical studies that comparatively assess predictive quality of disorder predictors [33,49–59]. These comparative studies include several community assessments, such as Critical Assessment of Structure Prediction (CASP) between CASP5 to CASP10 [53–58] and Critical Assessment of Intrinsic Protein Disorder (CAID) [52]. The community assessments involve evaluation of predictors on blind test datasets (i.e., datasets that were not available to the authors of the predictors) by independent assessors who do not take part in the competitions utilizing tests and metrics that are widely accepted by the community.

The predictive architectures used to develop disorder predictors are typically divided into three categories [42,43,46,47]: (1) sequence scoring functions; (2) machine learning models; and (3) meta-predictors. The first category uses additive and/or weighted functions, some of which are grounded in physical principles governing protein folding, to process the input protein sequence and sequence-derived structural and evolutionary information. Representative disorder predictors that fall into this category include FoldIndex [60], IU Pred [61,62], and IU Pred 3 [63]. The machine learning predictors apply models that are trained from data using a variety of machine learning algorithms, such as support vector machines [64–66], regression [67], conditional random fields [68–70], radial basis function networks [71], and shallow neural networks [36,72–76]. Example popular machine learning predictors include DisEMBL [36], DISOPRED [75,76], PONDOR [73], and PrDOS [64]. The meta-predictors use multiple disorder predictions as inputs to re-predict disorder. The underlying rationale was to exploit potential complementarity among the input disorder predictions to generate a new prediction that would improve over the inputs. These efforts were also fueled by the availability of diverse sequence-scoring and machine learning predictors and studies that empirically show that well-designed meta predictors indeed produce predictions that outperform their inputs [77–79]. Representative example meta-predictors of disorder include metaPrDOS [80], MFDp [65,81,82], CsPritz [83], disCoP [77,84], and MobiDB-lite [78]. We observe that some meta-predictors use machine learning algorithms (e.g., metaPrDOS [80] and MFDp [65]), which means that they can be cross-listed in both categories.

Results of CASP10, the most recent CASP community assessment that covers disorder prediction (i.e., subsequent CASP experiments do not include disorder predictions), reveal that the top three predictors belong to the machine learning (PrDOS and DISOPRED) and meta-predictor (MFDp) categories [58]. However, a recent survey notes a rapid influx of a new subfamily of machine learning methods that relies on deep neural networks (DNNs) after the first DNN-based method was released in 2013 [41]. DNNs differ from shallow neural networks, which were commonly used to implement disorder predictors in early 2000 s [36,72–76], by use of multiple hidden layers and more sophisticated types of neurons and connections. The shift to the deep network models is motivated by their favorable levels of predictive performance when compared with the other types of disorder predictors. In particular, we observe that the best performing methods from the just completed CAID experiment [85], which include fDpnn [86], SPOT-Disorder2 [87], RawMSA [88] and AU CPred [89], rely on DNNs. Motivated by their growing numbers and success, we provide the first review of the DNN-based disorder predictors. We identify and summarize 13 DNN-based disorder predictors that were developed since 2013. We analyze trends in the development of these predictors and empirically compare predictive quality produced by the deep learners against the other types of disorder predictors based on results produced on blind test dataset from the CAID experiment. We also comment on future prospects in the development of the DNN-based disorder predictors.

2. Prediction of intrinsic disorder using deep learning

Nowadays, deep learning is widely used to develop methods that predict protein structure and function. Perhaps the most obvious example is protein structure prediction where deep learning models, such as AlphaFold, have deservedly dominated over other types of methods [90–93]. Moreover, deep learning is utilized to predict other structural aspects of proteins, such as contacts [94], secondary structure [95] and torsional angles [96]. DNNs are also successfully applied to predict protein function [97–99], protein-drug interactions [100,101], and functional sites [102–104].

The intrinsic disorder prediction field was not immune to the infusion of the deep learning-based approaches. The first DNN-based disorder predictor, DNDisorder [105], was published in 2013. Table 1 summarizes a comprehensive list of 36 disorder predictors that were published since that time. This list contextualizes the efforts to develop deep learning predictors in a broader setting of the entire disorder prediction field. We identify the 36 predictors using a wide-ranging list of sources including databases of disorder predictions: MobiDB [122], D²P² [123] and DescribePROT [124]; community assessments and surveys that were published on or after 2013 [33,41–43,46,47,50,52,58,59], and a manual search of relevant articles from PubMed that we collect using the “(disorder[Ti tle]) AND (prediction[Ti tle]) AND protein” query. Table 1 reveals that 13 out of the 36 recent disorder predictors use deep learning models. We find that it took two more years for the second DNN-based predictor, DeepCNF-D, to be published in 2015 [112]. The following three years include similarly low numbers of new deep learning tools, with two methods published in 2016, one in 2017, and one more in 2018. Year 2019 marks a turning point in the efforts to develop DNN-based disorder predictors, with two tools published in 2019, two in 2020, and four in 2021. Fig. 1 conveniently summarizes the corresponding trends. It highlights the gradual shift to developing predictors that rely on deep networks and the fact that these methods constitute majority (58%) of the predictors that were published over the last three years (green line in Fig. 1). We also note that the consistent levels of the release of new methods that range between 11 and 13 per every three-years long interval.

Table 1 provides a few additional insights. We manually check websites of the corresponding methods and find that 23 out of 36 predictors (over 60%) are available to the end users as either standalone software (5 methods), webserver (10 methods) or in both modalities (10 methods). Interestingly, all DNN-based predic-
tors that were published after 2016, except for flDPlr, are among the publicly available tools. This rate of availability is substantially better compared to related areas including prediction of protein-binding and RNA-binding residues where the availability is around 40% [103,125]. The webservers are a convenient option to less programming savvy end users, such as some biochemists or structural biologists. In this case, predictions are performed on the webserver end and users are not required to install and run

Table 1
Summary of intrinsic disorder predictors that were developed since 2013 when the first deep learning-based method was released. The predictors are sorted in the chronological order of their year of publications. "*" denotes predictors that are used in Fig. 3.

| Predictor name                  | Year published | Reference | Applies DNN | Availability | URL                                                                 |
|---------------------------------|----------------|-----------|-------------|--------------|----------------------------------------------------------------------|
| MFDP2                           | 2013           | [81]      | No          | WS           | https://biomine.cs.vcu.edu/servers/MFDp2/                             |
| DNDisorder                      | 2013           | [105]     | Yes         | N/A          | N/A                                                                  |
| preDNDisorder                   | 2013           | [105]     | No          | N/A          | N/A                                                                  |
| Ulg-GIAGA                       | 2013           | [108]     | No          | N/A          | N/A                                                                  |
| DisMeta                         | 2014           | [107]     | No          | WS           | https://montelionelab.chem.rpi.edu/dismeta/                           |
| discoP                          | 2014           | [77,84]   | No          | WS           | https://biomine.cs.vcu.edu/servers/discoP/                           |
| DynaMine                        | 2014           | [67,108]  | No          | SP + WS      | https://dynamine.ibsquare.be/                                        |
| PON-Diso                        | 2014           | [109]     | No          | WS           | https://structure.bmce.lu.se/PON-Diso                                |
| DISOPRED3*                      | 2015           | [75]      | No          | SP + WS      | https://biomine.cs.vcu.edu/servers/disCop/                            |
| s2D-1                           | 2015           | [110]     | No          | N/A          |                                                                      |
| s2D-2*                          | 2015           | [110]     | No          | N/A          |                                                                      |
| DisoMCS                         | 2015           | [111]     | No          | N/A          |                                                                      |
| DeepCNF-D                       | 2015           | [112]     | Yes         | SP           | https://home.ittc.edu/~wangsheng/software.html                        |
| AUCpredD*                       | 2016           | [89]      | Yes         | N/A          | N/A                                                                  |
| AUCpredD-op*                    | 2016           | [89]      | Yes         | N/A          | N/A                                                                  |
| DisPredict (DisPredict2)*       | 2016           | [66]      | No          | SP           | https://github.com/tamjindul/DisPredict2_PSEE                       |
| MobidB-lite*                    | 2017           | [78]      | No          | WS           | https://mobidb.bio.unipd.it/                                         |
| SPOT-Disorder*                  | 2017           | [113]     | Yes         | SP + WS      | https://sparkslab.org/server/spot-disorder/                           |
| rUPred2A-long*                  | 2018           | [114]     | No          | SP + WS      | https://iupred2a.ite.hu/                                              |
| rUPred2A-short*                 | 2018           | [114]     | No          | SP + WS      | https://iupred2a.ite.hu/                                              |
| pyHCA*                          | 2018           | No        | No          | SP           | https://github.com/T-B-F-pyhCA/                                      |
| SPOT-Disorder-Single*           | 2018           | [115]     | Yes         | SP + WS      | https://sparkslab.org/server/spot-disorder-single/                    |
| Predictor by Zhao and Xue       | 2018           | [116]     | No          | N/A          |                                                                      |
| IDP-CRF                         | 2018           | [69]      | No          | N/A          |                                                                      |
| rawMSA*                         | 2019           | [88]      | Yes         | SP           | https://bitbucket.org/Clami66/rawmsa/src/master/                     |
| SPOT-Disorder2*                 | 2019           | [87]      | Yes         | SP + WS      | https://sparkslab.org/server/spot-disorder2/                          |
| Spark-IDPP                      | 2019           | [117]     | No          | No           | N/A                                                                  |
| IDP-FSP                         | 2019           | [70]      | No          | No           | N/A                                                                  |
| DisoMine*                       | 2020           | No        | Yes         | WS           | https://www.bio2byte.be/b2btools/disomine/                           |
| ODINPred                        | 2020           | [118]     | No          | WS           | https://protein.chem.au.dk/odinpred/                                 |
| IDP-Seq2Seq*                    | 2020           | [119]     | Yes         | WS           | https://bitlab.net/IDP-Seq2Seq/                                      |
| fIDPnn*                         | 2021           | [86]      | Yes         | SP + WS      | https://biomine.cs.vcu.edu/servers/fIDPnn/                            |
| fIDP*                           | 2021           | [86]      | No          | No           | N/A                                                                  |
| IUPred3                         | 2021           | [63]      | No          | SP + WS      | https://iupred2.ite.hu/                                               |
| RFPR-IDP*                       | 2021           | [120]     | Yes         | WS           | https://bitlab.net/RFPR-IDP/server                                    |
| Metapredict*                    | 2021           | [121]     | Yes         | SP + WS      | https://github.com/idptools/metapredict                              |

1 "No" means that a given predictor was not published in a peer-reviewed journal but was included based on participation in the CASP and/or CAID assessment.
2 Availability: released as "SP" (standalone program), "WS" (web server). "No" not released as either SP (standalone program) or WS (web server), and "N/A" (not available) SP and/or WS were released at the time of publication (i.e. URL was provided in the original article) but they were not available as of February 2022 when the access was tested.

Fig. 1. Development of disorder predictors since 2013 when the first deep learning-based predictor was released. The left/right y-axis gives the number/fraction of predictors in a given time period. The predictors are color-coded where green represents deep neural network-based methods and blue represents other types of predictors.
Table 2
Summary of intrinsic disorder predictors that use deep neural network models. The predictors are sorted in the chronological order of their year of publications. X marks inputs that are used by a given predictor. **"*" denotes predictors that are used in Fig. 3.

| Predictor name          | Year published | Inputs | Evolutionary features | Predicted structural feature | Physicochemical properties | Network architecture | Type5 | Size6 | AUC | Runtime7 |
|-------------------------|----------------|--------|-----------------------|-----------------------------|---------------------------|----------------------|-------|-------|-----|----------|
| DNdisorder              | 2013           | X X    | X                     | X                           | X                         | RBM                  | Moderately deep | N/A  | N/A  |
| DeepCNF-D               | 2015           | X X    | X                     | X                           | X                         | CNN                  | Moderately deep | N/A  | N/A  |
| AUCpredD                | 2016           | X X    | X                     | X                           | X                         | CNN                  | Moderately deep | 0.757 | 7.0   |
| AUCpredD-mp             | 2016           | X X    | X                     | X                           | X                         | CNN                  | Moderately deep | 0.751 | <0.5 |
| SPOT-Disorder*          | 2017           | X X    | X                     | X                           | X                         | BRNN                 | Moderately deep | 0.744 | 5.0   |
| SPOT-Disorder-Single*   | 2018           | X X    | X                     | X                           | X                         | BRNN + CNN           | Deep              | 0.757 | 0.8–1.0 |
| rawMSA*                 | 2019           | X X    | X                     | X                           | X                         | BRNN + CNN           | Very deep         | 0.780 | >10.0 |
| SPOT-Disorder2*         | 2019           | X X    | X                     | X                           | X                         | BRNN + CNN           | Very deep         | 0.760 | >10.0 |
| DisoMine                | 2020           | X X    | X                     | X                           | X                         | BRNN                 | Deep              | 0.765 | <0.5 |
| IDP-Seq2Seq*            | 2020           | X X    | X                     | X                           | X                         | FFNN                 | Moderately deep | 0.814 | 0.5–1.0 |
| flDPnn*                 | 2021           | X X    | X                     | X                           | X                         | BRNN + CNN           | Deep              | 0.722 | <0.5 |
| RFPR-IDP*               | 2021           | X X    | X                     | X                           | X                         | BRNN                 | Deep              | 0.746 | <0.5 |
| MetaPredict*            | 2021           | X      | X                     | X                           |                           | BRNN                 | Deep              |       |       |

1 The input sequence was encoded and directly used as predictive input.
2 Evolutional features computed from the input sequence including position-specific scoring matrix (PSSM), entropy-based conservation, and multiple sequence alignment.
3 Structural features predicted from the input sequence, such as secondary structure, solvent accessibility, and half-screw exposures.
4 Physicochemical properties of the amino acids in the input sequence including polarizability, hydrophobicity, and isoelectric point.
5 Type of the deep learning neural network used: “RBM” (Restricted Boltzmann Machine); “CNN” (Convolutional Neural Network); “BRNN” (Bidirectional Recurrent Neural Network); and “FFNN” (Feed Forward Neural Network).
6 The number of hidden layers: moderately deep with 2 to 3 layers; deep with 4 to 5 layers; and very deep with over 5 layers.
7 The average runtime in minutes to predict one amino acid sequence. N/A denotes that the results could not be collected since a working implementation of the corresponding predictor is not available.

The recently completed CAID experiment reveals that some of the DNN-based solutions provide favorable predictive performance when compared to other types of disorder predictors [52]. This conclusion is perhaps best captured with the following quote: “The SPOT-Disorder2 and flDPnn, followed by RawMSA and AUCpredD, are consistently good. However, flDPnn is at least an order of magnitude faster than its competitors, and it succeeded on all sequences, whereas SPOT-Disorder2 skipped 5% of sequences as a result of a length limitation.” [85]. While these four best predictors rely on deep learning, they implement the underlying predictive models using very different designs. More specifically, flDPnn relies on moderately deep CNN architecture [86], SPOT-Disorder2 and RawMSA are very deep hybrids of CNN and BRNN [87,88], while AUCpredD utilizes moderately deep CNN topology [89]. This observation suggests that accurate disorder prediction can be accomplished using different types of deep learners.

We provide a wider comparison of the predictive performance of deep learners. We cover 11 DNN-based methods that exclude only the two oldest methods, DNdisorder and DeepCNF-D. DNdisorder is not available to the end users (Table 1) while the standalone version of DeepCNF-D requires specific feature encoding of the sequence that we could not reproduce. We compare predictive performance of the remaining 11 deep learners using the annotated CAID dataset from https://idpcentral.org/caid/data/1/ and https://idpcentral.org/caid/data/1/reference/disprot-disorder.txt. This dataset includes 652 protein sequences and 337,908 amino acids, with 838 disordered regions and 54,820 disordered residues. For the 8 of the 11 predictors that were evaluated in CAID (i.e., AUCpred [89], AUCpred-mp [89], DisoMine [129], flDPnn [86], RawMSA [88], SPOT-Disorder [113], SPOT-Disorder-Single [115],
and SPOT-Disorder2 [87], we parse their CAID predictions from https://idpcentral.org/caid/data/1/predictions/. We collect results for the other three methods (IDP-Seq2Seq [119], RFPR-IDP [120], and Metapredict [121]) using the web servers and standalone programs provided by the authors. Table 2 shows that the predictive quality of deep learners measured with the area under the ROC curve (AUC) ranges between 0.722 for RFPR-IDP and 0.814 for fDPnn.

We further evaluate whether differences in the AUCs of the 11 predictors are robust across different datasets by comparing results across 20 randomly selected disjoint sets of 5% of proteins from the CAID dataset. We assess significance of differences in AUCs between the best-performing fDPnn and the other methods. We use the t-test if the underlying data are normal; otherwise, we use the Wilcoxon signed-rank test; we test normality with the Anderson-Darling test at the 0.05 significance. We find that fDPnn and RawMSA are not statistically different (p-value ≥ 0.05) but fDPnn is statistically better than the other 9 methods (p-value < 0.05). We similarly quantify significance of differences between RFPR-IDP that has the lowest AUC and the other 10 predictors. This analysis reveals that SPOT-Disorder, Metapredict, AUCpreD-np and IDP-Seq2Seq produce predictions that are not statistically better than RFPR-IDP (p-value ≥ 0.05). The remaining 4 predictors that include AUCpreD, SPOT-Disorder-Single, SPOT-Disorder2, and DisoMine are significantly better than RFPR-IDP (p-value < 0.05) and significantly worse than fDPnn (p-value < 0.05). Correspondingly, we identify 3 groups of the DNN-based predictors: 1) fDPnn and RawMSA that secure the best results (AUC > 0.78); AUCpreD, SPOT-Disorder-Single, SPOT-Disorder2, and DisoMine that obtain the second-best performance (0.755 < AUC < 0.78); and RFPR-IDP, SPOT-Disorder, Metapredict, AUCpreD-np and IDP-Seq2Seq that provide more modest levels of predictive quality (0.720 < AUC < 0.755).

We also analyze an average per-protein runtime for the predictors from Table 2. Similar to the analysis of the predictive performance, we could not perform this analysis for NDisorder and DeepCNF-D that do not provide working implementations. We extract the runtime data from the CAID results for the eight methods that participated in this experiment [52], and we estimate it for the other three methods (IDP-Seq2Seq, RFPR-IDP and Metapredict) based on the implementations provided by the authors. We find that the runtime of the 11 predictors varies widely (Table 2), with the fastest predictors that produce results in several seconds and the slowest that require over 10 min for the same task.

Using the above analysis, Fig. 2 compares the 11 available predictors based on three key characteristics: predictive performance quantified with AUC, speed measured with runtime, and mode of availability. The predictors are sorted in the chronological order of their year of publications. The color-coded scores represent quality where 2 (dark blue) is best, 1 (blue) is intermediate, and 0 (light blue) is worst. The AUC values are categorized into three groups using statistical test that measures robustness of differences between predictors over different protein sets; details are described in the text. Methods with AUCs that are not statistically different (p-value ≥ 0.05) from the best (worst) performing fDPnn (RFPR-IDP) are labeled with 2 (0), while the remaining predictors are labeled with 1. The runtime is divided into three ranges: < 1 min (score of 2); between 1 and 10 min (score of 1); and > 10 min (score of 0). The availability score counts the number of modes where 2 means that both SP (standalone program) or WS (web server) are available and 1 that either SP or WS are available.

3. Deep learning methods outperform other predictors of intrinsic disorder

Motivated by the finding that the top performing predictors in CAID are deep learners [52,85], we investigate whether this result can be extended more broadly to other DNN-based methods. More specifically, we compare the results for the 11 available deep learning-based disorder predictors from Table 2 against the results of other types of methods that we collect using the same CAID data. This analysis covers a comprehensive set of 29 disorder predictors including 11 deep learners that are annotated with * in Table 2 and 18 methods that use the other types of models. The latter group includes 12 machine learning predictors (DisEMBL-465 [36], DisEMBL-HL [36], DISOPRED3 [75], DisPredict2 [66], Espirit-D [130], Espirit-N [130], Espirit-X [130], IDPÀr [86], PONDY VSL2B [131], PreDisorder [74], RONN [132], and s2D-2 [110]); 5 sequence scoring function-based methods (FoldUnfold [133], IsUnstruct [134], IUPred2A-long [114], IUPred2A-short [114], and pyHCA [135]) and one meta-predictor (MobiDB-lite [78]). We mark these methods with * in Table 1, except for DisEMBL-465, DisEMBL-HL, JRONN, FoldUnfold, PONDY VSL2B, PreDisorder, IsUnstruct, Espirit-D, Espirit-N, and Espirit-X that were published before 2013. We quantify the predictive performance using four popular metrics that are consistent with the measures used in the most recent community assessments [52,58], including AUC, area under the precision-recall curve (AUPR), F1 and Matthews correlation coefficient (MCC). Finally, we quantify statistical significance of differences in the predictive performance between the results of the 11 deep learners and the 18 other methods. We test normality of the measured scores with the Anderson-Darling test and we apply the student t-test for normal data and the Wilcoxon test otherwise. Fig. 3 summarizes the corresponding empirical results. The median AUC of the deep learners is 0.76 vs. 0.73 for the other tools. We observe similarly substantial magnitude of differences for the other metrics, with median AUPR of 0.35 vs. 0.31, median F1 of 0.42 vs. 0.39 and median MCC of 0.29 vs. 0.26. The statistical analysis reveals that the DNN-based methods outperform the other disorder predictors by a statistically significant margin across the four metrics (p-value < 0.05). This consistent and statistically significant trend suggests that the deep neural networks are the best choice to develop accurate disorder predictors.
4. Summary and outlook

Disorder prediction is an active and well-establish research area with over 40 years of history. The first DNN-based disorder predictor was published in 2013 and 12 more deep learners were published since. We find that majority of the disorder predictors that were developed in the last three years utilize deep neural networks. The popularity of this design is motivated by several factors. First, these models can be molded into many different architectures that are flexible to use diverse types of inputs. Our analysis of the 13 DNN-based disorder predictors reveals that they rely on very diverse designs that explore different inputs, topologies and sizes. Second, our empirical results reveal that the DNN-based predictors are in general statistically better when directly compared against a representative collection of the other types of predictive models. This conclusion is in line with the results of the recent CAID experiment where the top four predictors are deep learners [52,85]. Third, our multifaceted comparison of the deep learners provides useful clues for the end users by identifying methods that are accurate, fast and widely available. We identify several well-rounded predictors that include flDPnn (very accurate, very fast, and available in multiple ways), SPOT-Disorder-Single (accurate, very fast, and available in multiple ways), DisoMine (accurate and very fast) and Metapredict (very fast and available in multiple ways). These results and accolades support conclusions of the recent article that say “deep-learning-based methods will likely continue to show the greatest potential for future improvement” [85].

Our analysis finds that the architectures of the current deep learners are considerably diverse. This suggests that the optimal architecture is yet to be identified. We reason that this should be a hybrid design to accommodate for the underlying variety of different types/flavors of disorder [136–138]. For instance, IDRs cover a wide spectrum of sizes, from short regions that are frequently localized at the sequence termini to very long regions that span the entire protein sequence [139,140]. IDRs also vary in their conformational space, which is signified by their classification into the native coils, native pre–molten globules and native molten globules [4,141]. Moreover, IDRs carry out many different functions, and some of them are multifunctional (moonlighting) [142,143], which results in many different biases in their sequences [4,137]. Interestingly, design of the recently published and well-rounded flDPnn suggests that predictive quality can be improved by innovating inputs that are fed into the deep networks [86]. The authors point to multiple options including development of extended sequences profiles that cover relevant sequence-derived protein characteristics beyond the commonly-used inputs listed in Table 2, and construction of aggregate features that quantify sequence bias at the region or whole sequence level. These two future directions go hand in hand given the fact that the hybrid deep learners are inherently capable of handling diverse and large inputs.

While most of recently released predictors of intrinsic disorder utilize DNNs, this is not necessarily the case for the methods that predict binding IDRs. There are close to 20 predictors of disordered protein-binding regions [144] and several methods that predict IDRs that interact with nucleic acids and lipids [42,145]. Examples of the recently published tools include FLIPPER [146], SPOT-MoRF [147], OPAL+ [148], DisoLipPred [149] and DeepDISObind [150]. The CAID experiment evaluated close to a dozen of these predictors and concluded that “disordered binding regions remain hard to predict” [52], motivating further efforts in this area. One of the potential reasons for the low predictive performance of these tools is a relatively low utilization of the deep learning architectures. We identify only a handful of DNN-based predictors of binding IDRs including SPOT-MoRF [147], MorRPred_en [151], en_DCNNMoRF [152], DeepDISObind [150], and DisoLipPred [149]. A similar situation is true in the context of prediction of disordered linker regions where neither of the two currently available methods, DFLpred [153] and APOD [154], applies deep learning and their predictive performance is relatively limited. Given the success of DNNs in the disorder prediction, we believe that this technology could be successfully applied to strengthen the quality of the predictors of binding IDRs and disordered linkers.

Funding

This research was funded in part by the National Science Foundation (grant 2125218) and the Robert J. Mattauch Endowment funds to LK.

CRediT authorship contribution statement

Bi Zhao: Formal analysis, Data curation, Investigation, Validation, Writing – original draft, Writing – review & editing.
Lukasz Kurgan: Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
B. Zhao and L. Kurgan

Computational and Structural Biotechnology Journal 20 (2022) 1286–1294

[67] Cilia E, Pancra R, Tompa P, et al. From protein sequence to dynamics and disorder with DynaMine. Nat Commun 2013;4:2741.
[68] Wang L, Sauer UH. On-D-CRF: predicting order and disorder in proteins using [corrected] conditional random fields. Bioinformatics 2008;24(11):1401–2.
[69] Liu YM, Wang XL, Liu B. IDP-CRF: Intrinsically Disordered Protein/Region Identification Based on Conditional Random Fields. International Journal of Molecular Sciences. 2018;Sept 19:9.
[70] Liu Y, Chen S, Wang X, et al. Identification of intrinsically disordered proteins and regions by length-dependent predictors based on conditional random fields. Mol Ther Nucleic Acids 2016;5:396–404.
[71] Su CT, Chen CY, Hsu CM. iPDA: integrated protein disorder analyzer. Nucleic Acids Res 2007;35:W465–72.
[72] Xue B, Dunbrack RL, Williams RW, et al. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010;1804(4):996–1010.
[73] Peng K, Vucetic S, Radijovic P, et al. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 2015;13(3):35–60.
[74] Deng X, Eickholz J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinform 2009;21(10):436.
[75] Jones DT, Czezott D. DISOPRED3: precise disorder region predictions with annotated protein-binding activity. Bioinformatics 2015;31(11):857–63.
[76] Jones DT, Ward JJ. Prediction of disordered regions in proteins from position specific score matrices. Proteins 2003;53(Suppl 6):573–8.
[77] Fan X, Kurgan L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J Biomol Struct Dyn 2014;32(3):448–64.
[78] Necci M, Piovesan D, Dosztanyi Z, et al. MobiDB-lite: fast and highly specific assessment of RNA-binding residue predictions with analysis by RNA type. Nucleic Acids Res 2020;36(18):4729–38.
[79] Wang K, Hu G, ZW, et al. Comprehensive survey and comparative assessment of RNA-binding residue predictions with analysis by RNA type. Nucleic Acids Res 2020;36(18):4729–38.
[80] Wang K, Chen J, Cheng J. CASPR: predicting protein disorder using boosting and deep networks. BMC Biol 2013;11(4):88.
[81] Becker J, Maes F, Wehenkel L. On the encoding of proteins for disordered regions prediction. PLoS ONE 2013;8(12):e82252.
[82] Wang X, Aeron TB, Montelione GT. DisMeta: a meta server for construct design and optimization. Methods Mol. Biol. (Clifton, NJ) 2014;1091:1–16.
[83] Cilia E, Pancra R, Tompa P, et al. The DynaMine webserver: predicting protein dynamics from sequence. Nucleic Acids Res. 2014 Jul 42(Web Server issue):W264-70.
[84] Ali H, Urolagin S, Gurarslan O, et al. Performance of protein disorder prediction programs on amino acid substitutions. Hum Mutat 2014;35(7):794–804.
[85] Segalman R, Camilloni C, Fariselli P, et al. The 2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. J Mol Biol 2015;427(4):582–96.
[86] Wang Z, Yang Qi, Li T, et al. DISOMC3: accurately predicting protein intrinsically disordered regions using a multi-class conservative score approach. PLoS ONE 2015;10(6):e0128334.
[87] Wang S, Wang S, Ma J, et al. DeepCNN-D: predicting protein order/disorder regions by weighted deep convolutional neural fields. Int J Mol Sci 2015;16(8):17315–30.
[88] Hanson J, Yang Y, Palivai K, et al. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neuronal networks. Bioinformatics 2017;33(5):895–902.
[89] Tenenbaum BJ, Erds G, Dosztanyi Z, UPPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 2018;46(W1):W329–37.
[90] Hanson J, Palivai K, Kou Y. Accurate single-sequence prediction of protein intrinsic disorder by an ensemble of deep recurrent and convolutional architectures. J Chem Inf Model 2015;58(11):2369–76.
[91] Zhao B, Xue B. Decision-tree based meta-strategy improved accuracy of disorder prediction and identified novel disordered residues inside binding motifs. Int J Mol Sci 2018;19(10).
[92] Malyshak-Mrozek B, Baron T, Mrozek D. Spark-IDPP: high-throughput and scalable prediction of intrinsically disordered protein regions with Spark and Apache on the Cloud. Cluster Comput 2019;22(2):487–508.
[93] Dass R, Mulder FAA, Nielsen JT. ODIOPred: comprehensive prediction of protein order and disorder. Sci Rep 2020;10(1):14780.
[94] Tang Y, Pang YH, Liu B. IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning. Bioinformatics 2021;36(21):W329–37.
[95] Emenecker RJ, Griffith D, Holehouse AS. Metapredit: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biochips J 2021;35:20(3):4312–9.
[96] Provesan D, Necci M, Escobedo N, et al. MobIDB: intrinsically disordered proteins in 2021. Nucleic Acids Res 2021;49(D1):D361–7.
[97] Raes J, Euretig C, Aymerich M. Long-range interactions and protein evolution. Nucleic Acids Res 2014;42(8):3453–66.
[98] van der Geer J, Hanraads JAJ, Lwwters R. The art of writing a scientific paper. J J L 2000;30(2):25–29.
[99] Wittig M, van Eekelen M, van Eekelen M, et al. DeepIDP: accurate prediction of protein disorder with deep learning. Bioinformatics 2018;34(4):460–8.
[100] Muller C, Rabal O, Diaz GC. Artificial intelligence, machine learning, and deep learning in real-life drug design cases. Methods Mol Biol 2021;2390:383–407.
Yang ZR, Thomson R, McNeil P, et al. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005;21(16):3369–76.

Galitskaya OV, Carbuzymsky SO, Lobanov MY. FoldITUnfold: web server for the prediction of disordered regions in protein chain. Bioinformatics 2006;22(23):2948–9.

Lobanov MY, Galitskaya OV. The Ising model for prediction of disordered residues from protein sequence alone. Phys Biol 2011;8(3):035004.

Bitard-Feildel T, Callebaut I. HCAtk and pyHCA: A Toolkit and Python API for the Hydrophobic Cluster Analysis of Protein Sequences. bioRxiv. 2018:249995.

Necci M, Piovesan D, Tosatto SC. Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. Protein Sci 2016;25(12):2164–74.

Desana A, Forcelloni S, Porrello A, et al. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. PLoS ONE 2019;14(8):e0217889.

Howell M, Green R, Kileen A, et al. Not that rigid midgets and not so flexible giants: on the abundance and roles of intrinsic disorder in short and long proteins. J Biol Syst 2012;20(4):471–511.

Uversky VN. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. FEBS Lett 2013;587(13):1891–901.

Nielsen JT, Mulder FA. There is diversity in disorder—“in all chaos there is a cosmos, in all disorder a secret order”. Front Mol Biosci 2016;3:4.

Uversky VN. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta 2013;1834(5):932–51.

Meng F, Kurgan L. High-throughput prediction of disordered moonlighting regions in protein sequences. Proteins 2018;86(10):1097–110.

Shchankov NN, Bustos DM. Intrinsic disorder associated with 14-3-3 proteins and their partners. Prog Mol Biol Transl Sci 2019;166:19–61.

Katuwawala A, Peng Z, Yang J, et al. Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions. Comput Struct Biotechnol J 2019;17:454–62.

Katuwawala A, Ghadermarzi S, Kurgan L. Computational prediction of functions of intrinsically disordered regions. Prog Mol Biol Transl Sci 2019;166:341–69.

Meng F, Kurgan L. DFLpred: High-throughput prediction of disordered lipid binding residues in protein sequences with deep recurrent networks and transfer learning. Bioinformatics 2020;36(4):1107–13.

Fang C, Moriwaki Y, Li C, et al. MoRFPred_en: Sequence-based prediction of MoRFs using an ensemble learning strategy. J Bioinform Comput Biol 2019;17 (6):1940015.

Fang C, Moriwaki Y, Tian A, et al. Identifying short disorder-to-order binding regions in disordered proteins with a deep convolutional neural network method. J Bioinform Comput Biol 2019;17(1):1950004.

Meng F, Kurgan L. DFUnpred: High-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics 2016;32(12):i341–50.

Peng Z, Xing Q, Kurgan L. APoD: accurate sequence-based predictor of disordered flexible linkers. Bioinformatics 2020;36(Supplement_2):i754–61.