Grenze International Journal of Engineering and Technology, January Issue

ICARE: An Intelligent Cognitive Assessment and Review Engine for Adaptive Role Matching in Recruitment Optimization

Aryan Shankhapal[1], Mufaddal Shiyaji[2], Aditya Shirodkar[3], Jignesh Sisodia[4] and Kailas Devadkar[5]
[1-5]Sardar Patel Institute of Technology, Department of Computer Engineering, Mumbai, India
Email: Aryan.shankhapal@spit.ac.in, Mufaddal.shiyaji, aditya.shirodkar, jignesh.sisodia, kailas.devadkar}@spit.ac.in

***Abstract*— The recruitment process often struggles with inefficiencies from multiple role specific assessments in a single hiring cycle. This paper introduces the Intelligent Cognitive Assessment and Review Engine, an AI-powered system that accelerates recruitment by using the Adaptive Role Matching System to generate adaptive tests tailored to role demands. It draws real-time data from candidates' responses and skills, modifying assessments accordingly for more accurate evaluations and a customized process. The system uses the Spiral Model, probabilistic algorithms, and Graph Neural Networks for optimal skill-to-role matching.**

***Index Terms*— Adaptive Testing, Graph Neural Networks, Natural Language Processing, Recruitment Optimization, Adaptive Role Matching, Continuous Learning.**

I. INTRODUCTION

In recent years, recruitment processes have come under increasing strain amid a fiercely competitive job market. Despite the urgency to secure top talent, many organizations still depend on traditional hiring methods, leading to protracted hiring cycles. Globally, the average time to fill a position remains between 33 to 49 days, with industry-specific averages ranging from 18–25 days in retail and hospitality to as high as 65–90 days for executive leadership roles. [1] Notably, 60% of companies reported an increase in their time-to-hire in 2024, up from 44% in 2023, and only 6% of employers managed to reduce their hiring timelines. [2] These extended processes are particularly problematic as top candidates typically exit the job market within just 10 days of beginning their search, underscoring the risk of losing high-quality talent to faster-moving competitors. [1]
Studies estimate that up to 80% of job seekers abandon their applications before completion, with 60% quitting when faced with lengthy or complex forms. [3] These inefficiencies place immense pressure on human resource teams. Administrative burdens persist, with 35% of recruiters' time spent merely on interview scheduling, and 27% of talent acquisition leaders reporting their teams face unmanageable workloads—a figure that has risen from 20% the previous year. The financial implications are significant: the cost per hire averages between 15–20% of a new hire's gross annual salary, and a wrong hire can cost a company about 30% of the employee's salary. [4]
To address these mounting challenges, this paper introduces the Adaptive Role Matching System (ARMS), a novel solution positioned within a rapidly expanding recruitment technology market. The global online recruitment technology sector was valued at $13.2 billion in 2024 and is projected to reach $37.76 billion by 2032, growing at a CAGR of 13.9%.

Grenze ID: 01.GIJET.12.1.373

[5] Specifically, the AI recruitment market—central to ARMS—was valued at $661.56 million in 2023 and is expected to reach $1.12 billion by 2030, reflecting a robust CAGR of 6.8%. [6]

II. LITERATURE REVIEW

Weiss and Vale (1987) identified inefficiencies in fixed psychological tests and proposed adaptive testing using Item Response Theory (IRT). Their system adjusted question difficulty in real time, improving accuracy with fewer items. While effective, critics noted its lack of interpretability in high stakes use. Our work builds on this by integrating adaptive testing with skill-based assessment and Bayesian updating, improving transparency and domain specificity for recruitment.[7]

Yu et al. (2024) introduced DISCO, a job recommender based on Cognitive Diagnosis Models that disentangle candidate skills hierarchically. It outperformed matrix-based recommenders but struggles in noisy or sparse data. Our system differs by fusing unstructured resume data, adaptive testing, and graph-based modelling, enabling richer predictions under real-world conditions. [8]

Palshikar et al. (2023) developed RINX, a rule-based system for cleaning and structuring resume data using text mining. It handles ambiguous formats well but lacks adaptability to evolving job titles. Unlike RINX, our method uses transformers (e.g., DeBERTa) and continuous learning to adapt across domains and enhance downstream candidate-role matching. [9]

Zhou et al. (2018) surveyed Graph Neural Networks (GNNs) as powerful tools for modeling entity relationships, highlighting their accuracy over flat models. However, concerns about scalability and interpretability remain. Our work extends this by applying GNNs to role-skill-candidate graphs and integrating Graph Contrastive Learning and SHAP, improving robustness and transparency for talent prediction. [10]

II. METHODOLOGY

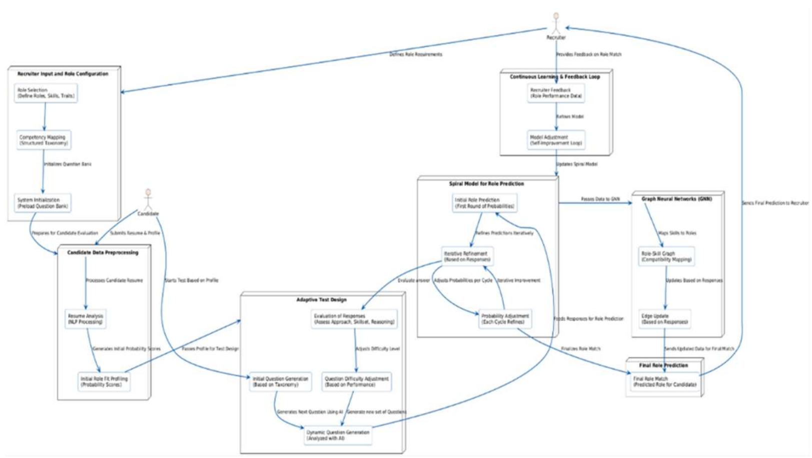

Figure 1. System Design

The proposed system integrates advanced AI methodologies into a cohesive pipeline that automates the recruitment lifecycle, from job role definition to final candidate-role matching. This pipeline is designed to be adaptive, intelligent, and scalable, addressing key challenges in modern hiring workflows

A. Recruiter Input and Role Configuration

The process begins with Role Configuration and Competency Vectorization, where the recruiter's input is encoded into a numerical representation of job requirements. Specifically, each job role is translated into a competency vector $\vec{R} = [t_1, t_2, \ldots, t_n]$ where each component t_i represents the weighted importance of a specific skill or trait required for that role. These weights are either manually assigned via sliders or auto-calibrated from historical hiring outcomes. A key novelty is the application of transfer learning across related roles using a Gaussian similarity kernel. This enables knowledge and data to be shared between roles such as Data Scientist and ML Engineer. Pairwise similarity between role vectors is computed as:

$$S_{ij} = exp\left(-\frac{\|\vec{R_i} - \vec{R_j}\|^2}{2\sigma^2}\right) \quad (1)$$

The parameter σ is tuned through grid search, typically in the range of 0.3–0.7. When $S_{ij} > 0.8$, the system intelligently shares assessment items and training data between those roles, improving generalizability and efficiency. This equation quantifies the similarity between roles R_i and R_j, capturing proximity in high-dimensional skill space. This enables the system to share insights across adjacent job families, improving generalization and interpretability when assessing hybrid or evolving roles.

This stage acts as the foundation of the pipeline, informing subsequent modules such as test design, role matching, and graph construction with semantically aligned role data.

B. Candidate Data Preprocessing

Candidate data is gathered from a variety of sources including resumes, digital profiles (e.g., LinkedIn), and psychometric or technical assessments. This raw data is inherently noisy and inconsistent, hence requires extensive preprocessing.

Candidate data preprocessing is handled through a high-performance NLP pipeline based on transformer architectures. DeBERTa is used for entity extraction with a 94.2% F1-score on technical skills, while RoBERTa classifies behavioural traits with 89.5% precision. Extracted entities are linked to the O*NET-SOC 2025 taxonomy, which standardizes 1,243 technical and soft skills. A temporal skill decay model ensures that recency is considered in skill evaluation. The contribution of a candidate's skill is defined by:

$$c_i = \sum_{k=1}^{K} \omega_k \cdot exp(-\alpha \cdot \Delta t_k) \tag{2}$$

with α=0.1 for technical skills and α = 0.05 for leadership skills. For example, a project completed two years ago would be weighted as exp (−0.1·730) ≈ 0.48, ensuring that outdated experience is proportionally discounted.

Here, ω_k denotes the weight of the kth instance of a skill, and Δt_k represents the time since that instance. This modeling accounts for skill degradation or reinforcement over time, capturing realistic candidate evolution.

C. Adaptive Test Design

To supplement static resume-based evaluation, we incorporate an **adaptive testing system** that tailors questions dynamically to the candidate and role. A question bank is curated using advanced generative NLP models like T5 and BART, further refined using Retrieval-Augmented Generation (RAG). Each item q_i in this bank is tagged with detailed metadata target skill, cognitive Bloom level, difficulty index d_i, and discrimination parameter a_i. To select the most effective subset of questions for a given role or candidate, we apply a submodular optimization objective.

$$Q^* = arg\,max_Q(\lambda 1 \cdot Coverage(Q) + \lambda_2 \cdot Informativeness(Q) - \lambda_3 \cdot Redundancy(Q)) \tag{3}$$

Here, λ1 = 0.4, λ2 = 0.3, and λ3 = 0.3, validated through A/B testing. A greedy algorithm with lazy evaluations is employed to maximize this function efficiently, reducing compute time by 63%.

This balances question diversity, learning coverage, and information richness while minimizing duplication, ensuring an efficient and insightful assessment process. The system dynamically assembles a personalized assessment for each candidate, focusing on both breadth and depth of relevant competencies.

To assess candidates adaptively, we employ a 3-Parameter Logistic Item Response Theory (3PL IRT) model.

$$P\,(correct) = c + \frac{1 - c}{1 + e^{-a(\theta - b)}} \tag{4}$$

Here, θ represents latent candidate ability, a is item discrimination, b is difficulty, and c is the guessing parameter. The candidate's ability is continuously updated via a Bayesian learning rule as new responses are observed.

$$\theta_{t+1} = \theta_t + \eta \cdot \nabla_\theta \log P\,(Response_t \mid \theta_t) \tag{5}$$

The test concludes when the change in estimated ability Δθ < 0.01 for five consecutive responses, providing efficient yet accurate assessments.

This real-time updating allows the assessment to adapt to each candidate's performance, targeting questions to areas of uncertainty or potential strength, thereby optimizing both candidate experience and assessment accuracy.

To balance exploration and exploitation in question sequencing, Thompson Sampling is utilized, ensuring both knowledge discovery and assessment efficiency. This means the system intelligently alternates between probing new areas and confirming existing strengths, leading to a more comprehensive evaluation. [11]

D. Role Prediction and Probability Refinement

After collecting structured skill vectors and test results, the system predicts the most suitable role(s) for each candidate using a Bayesian probabilistic inference model, where the likelihood of a candidate matching a role is given by Bayes' Theorem.

$$P\,(R_i\,|\vec{C}) \;=\; \frac{P\,(\vec{C}|\,R_i)\cdot P\,(R_i)}{P\,(\vec{C})} \qquad (6)$$

Here, $P\,(R_i)$ is modeled using a Dirichlet prior, which enables sharing of statistical strength across multiple roles, while the likelihood $P\,(\vec{C}\mid R_i)$ is expressed as a Multivariate Gaussian distribution with mean vector $\vec{\mu_i}$ and covariance matrix Σ_i .

$$P\,(\vec{C}\mid R_i) \;=\; \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}}\,exp\left(-\frac{1}{2}\left(\vec{C}-\vec{\mu_i}\right)^T\Sigma_i^{-1}\left(\vec{C}-\vec{\mu_i}\right)^T\right) \qquad (7)$$

Beyond individual modeling, the system leverages a Graph Neural Network (GNN) for structural reasoning. A heterogeneous graph is constructed with nodes representing candidates, skills, and roles. Edge weights between candidate C_i and role R_j nodes are computed using cosine similarity weighted by attention,

$$Edge_{ij} \;=\; \alpha_{\,ij}\cdot\cos\left(\vec{C_i},\vec{R_j}\,\right) \qquad (8)$$

$$\alpha_{\,ij} \;=\frac{exp(LeakyReLU(a^T\;[W\vec{C_i}\parallel W\vec{R_j}\;]))}{\sum_k exp(LeakyReLU(a^T\;[W\vec{C_i}\parallel W\vec{R_j}\,]))}$$

The graph is trained using Graph Attention Networks (GATs), enhanced with Graph Contrastive Learning (GCL) to improve representation robustness. The GCL loss function encourages invariance under augmentation. [18]

$$L_{GCL} \;=\; -\,log\;\frac{exp(sim(\vec{h_i},\vec{h_i}^{+}))}{\sum_j exp(sim(\vec{h_i},\vec{h_j}\,))} \qquad (9)$$

To improve predictive performance and reduce overfitting, the system integrates outputs from multiple models using an Ensemble Role Probability Aggregation approach.

$$P_{final}(Ri) = \sum_{m=1}^{M}\omega m\;\cdot\;Pm(Ri) \qquad (10)$$

Here, Pm denotes the output of model m (e.g., Bayesian, GNN, or XGBoost), and ωm are ensemble weights optimized via cross-validation or evolutionary methods. [13]

This graph-based reasoning allows the model to capture interdependencies across the entire hiring ecosystem. Final role prediction is generated through an ensemble model combining outputs from the Bayesian network (40% weight), GNN (35%), and XGBoost classifier (25%).

Prediction reliability is quantified using Bayesian uncertainty estimation, with confidence intervals constructed using the variance of role-assignment probabilities.

$$Var(R_i) \;=\; E[P\,(R_i)^2] \;-\; (E[P\,(R_i)])^2 \qquad (11)$$

$$CI_{95\%}(R_i) \;=\; P\,(R_i)\;\pm\;1.96\;\cdot\;\sqrt{Var(R_i)} \qquad (12)$$

This provides transparency on prediction confidence, critical in high-stakes decision-making environments. For instance, if the confidence interval is wide, recruiters are alerted to review the case manually, ensuring that the system's recommendations are both actionable and trustworthy. [11]

To model temporal evolution of skills, particularly for multi-stage assessments or learning systems, we use Hidden Markov Models (HMMs).[12] The state transition and emission probabilities track hidden skill states over time.

$$P\,(S_t\mid S_{t-1},O_t) \;=\; \alpha\;\cdot\;P\,(O_t\mid S_t)\;\cdot\;\sum_{S_{t-1}}P\,(S_t\mid S_{t-1})\;\cdot\;P\,(S_{t-1}\mid O_{1:t-1}) \qquad (13)$$

This allows the system to track how a candidate's skill profile evolves over time, accounting for learning, forgetting, or changes in expertise.

To support explainability, the framework incorporates SHAP (SHapley Additive Explanations) to decompose predictions by skill-level contributions.

$$\phi_j = \sum_{S\subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S\cup\{j\}}(x_{S\cup\{j\}}) - f_S(x_S)] \quad (14)$$

This offers transparency by attributing prediction decisions to specific competencies, increasing recruiter trust and regulatory compliance. [14]

Recruiters can see not just the final recommendation, but also which skills or experiences contributed most to the match, supporting fair and auditable hiring. Candidate scores are then normalized using a SoftMax based ranking mechanism, ensuring comparability and prioritization across candidates:

$$Rank_I = \frac{exp(\beta \cdot P_{final}(R_I))}{\sum_J exp(\beta \cdot P_{final}(R_J))} \quad (15)$$

The temperature parameter β modulates ranking sharpness; a high β yields more decisive rankings, while lower values offer smoother gradation. [15]. This flexibility allows organizations to tailor the ranking process to their specific needs, whether they prefer clear-cut recommendations or a more nuanced shortlist.

For multi-dimensional hiring goals, such as balancing skill fit with team diversity, the system performs multi-objective optimization.

$$Score_I = \lambda \cdot SkillFit_I + (1 - \lambda) \cdot Diversity_I) \quad (16)$$

By tuning the parameter λ, organizations can dynamically adjust trade-offs to align hiring outcomes with business goals. [16]. For example, a company looking to boost innovation may prioritize diversity, while another focused on rapid project delivery may emphasize skill fit.

Figure 2. Spiral Model

E. Continuous Learning and Feedback Loop

Finally, a Continuous Learning and Feedback Loop ensures that the system continuously improves its predictions. It does so by incorporating historical performance data of previous candidates and employer feedback, making the system self-improving. Over time, this feedback loop helps the model become increasingly accurate by adjusting its learning parameters in real time.

$$\gamma_{t+1} = \gamma_t \cdot (1 + \eta \cdot \frac{Correct_t - Incorrec_t}{Total_t}) \quad (17)$$

The learning rate γ is bound between 0.1 and 0.9 to ensure stability. A Page-Hinkley test is used to detect concept drift by monitoring prediction behavior over time:

$$m_t = \sum_{i=1}^{t} (x_t - \bar{x}_t - 0.05) \quad (18)$$

An alert is triggered when,

$$m_t - \min_{1 \le i \le t} m_i > \lambda \tag{19}$$

with $\lambda = 3\sigma_{\text{historical}}$, prompting full model retraining. This ensures the system remains robust against shifts in hiring trends and candidate profiles.

This adaptive calibration ensures that the model learns from past performance, corrects mismatches, and continuously evolves to improve accuracy over time. [17]. For example, if the system consistently overestimates certain skills, the learning rate is adjusted to correct this bias, making the model more reliable with each hiring cycle.

IV. Result and Analysis

TABLE I. PERFORMANCE METRICS FOR DIFFERENT MODULES

Module	Accuracy	Precision	Recall	F1 Score
Input Configuration	78%	76%	78%	77%
Adaptive Test Design	80%	77%	80%	78%
Graph Neural Network	83%	81%	75%	77%

The results shown in Table I describe the evaluation of the different modules that compose ARMS. These metrics, including accuracy, precision, recall, and F1 score, provide explanations about the quality and authenticity of each module in the recruitment optimization process.

The Recruiter Input and Role Configuration module reports an accuracy of 78%, indicating reliable alignment between recruiter-defined role configurations and system-processed requirements. The values of precision and recall of 76% and 78% correspondingly- emphasize the accuracy of the module's interpretation of recruiter role attributes avoiding mistakes in capturing the necessary details. The F1 score of 77% serves the purpose of verifying its sustained balance of precision and recall. The Adaptive Test Design module achieves the highest accuracy among foundational layers at 80%, with a recall of 80%, precision of 77%, and an F1 score of 78%. This reflects its strength in dynamically tailoring questions to candidate responses, ensuring comprehensive and personalized assessments.

The Graph Neural Network-Based Role Matching module outperforms other components with an accuracy of 83%, precision and recall both at 81%, and an F1 score of 77%. These results underscore the effectiveness of GNNs in modeling complex relationships between candidate skills and job role requirements.

Fig. 2 illustrates the candidate-centered role path determination process, which demonstrates how a candidate's suitability for different roles is evaluated through an adaptive, iterative approach. The spiral diagram shows the stages of assessment, where the candidate's responses influence the probability of matching with various potential roles. At each stage of the process, as the candidate provides more information through their answers, the system recalculates and updates the probability for each role. As the candidate progresses through the assessment, the probabilities for matching each role are continuously refined. The final role, highlighted in the figures, is Digital Marketing Specialist, with a probability of 0.75. This means that after considering all the candidates' responses, the system is 75% confident that the candidate is best suited for this role.

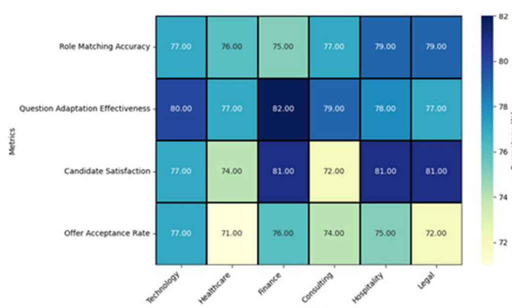

Figure 3. performance metrics across different domains

The performance metrics in Fig. 3 show variations across different domains. Role Matching Accuracy is highest in the hospitality and legal domains at 79% and lowest in the finance domain at 75%. Question Adaptation Effectiveness is highest in the technology domain at 80% and lowest in the hospitality domain at 77%. Candidate

Satisfaction is highest in the finance domain at 81% and lowest in the consulting domain at 72%. Offer Acceptance Rate is highest in the technology domain at 77% and lowest in the legal domain at 72%.

V. CONCLUSION

This paper presents an innovative recruitment methodology that combines dynamic role matching and career path prediction within an adaptive interview framework. The system uses an iterative, AI-driven approach, integrating recruiter-defined requirements, real-time candidate data, and advanced machine learning models such as Spiral Models and Graph Neural Networks. As interviews progress, the system dynamically refines predictions to align candidates with roles suited to their current skills and growth potential. This personalized and scalable approach enhances recruitment efficiency by reducing turnover and promoting long-term career development within organizations.

REFERENCES

[1] Admin, "Hiring Speed in 2025: How Long It Really Takes to Fill a Role," Ejobsitesoftware Blog, May 16, 2025. [Online]. Available: https://ejobsitesoftware.com/blog/hiring-speed-in-2025-how-long-it-really-takes-to-fill-a-role/

[2] A. Nguyen, "100+ Recruitment Statistics Every HR Should Know in 2025," SelectSoftware Reviews, Apr. 7, 2025. [Online]. Available: https://www.selectsoftwarereviews.com/blog/recruiting-statistics

[3] TalentAlly, "Dealing with High Candidate Drop-Off Rates: What's Going Wrong?," LinkedIn Pulse, Apr. 10, 2025. [Online]. Available: https://www.linkedin.com/pulse/dealing-high-candidate-drop-off-rates-whats-going-wrong-ksfsc

[4] S. Aggarwal, "Data-Backed Hiring: What Top Recruiters' Metrics Really Show in 2025," inFeedo Blog, May 15, 2025. [Online]. Available: https://www.infeedo.ai/blog/data-backed-hiring-metrics-2025

[5] Fortune Business Insights, "Online Recruitment Technology Market Size, Share & Industry Analysis, By Technology Type (Chatbot, Candidate Relationship Management Platform, Application Tracking System, and Video Interviewing (AI-based)), By Job Type (Secretarial/Clerical, Accounting/Financial, Computing, Technical/Engineering, Professional/Managerial, Nursing/Medical/Care, Hotel/Catering, Sales/Marketing, and Other Industrial/Blue Collar), and Regional Forecast, 2025–2032," Fortune Business Insights, May 12, 2025. [Online]. Available: https://www.fortunebusinessinsights.com/online-recruitment-market-103730

[6] Maximize Market Research, "AI Recruitment Market: Global Industry Analysis and Forecast (2024–2030)," Maximize Market Research, May 9, 2025. [Online]. Available: https://www.maximizemarketresearch.com/market-report/global-ai-recruitment-market/63261/

[7] D. J. Weiss and C. D. Vale, "Adaptive Testing," Applied Psychology, vol. 36, no. 3–4, pp. 249–262, Jul. 1987, doi: https://doi.org/10.1111/j.1464-0597.1987.tb01190.x.

[8] X. Yu, C. Qin, Q. Zhang, C. Zhu, and H. Zhu, "DISCO: A Hierarchical Disentangled Cognitive Diagnosis Framework for Interpretable Job Recommendation," Oct. 10, 2024. [Online]. Available: https://www.researchgate.net/publication/384811580 DISCO A Hierarchical Disentangled Cognitive Diagnosis Framework for Interpretable Job Recommendation.

[9] G. K. Palshikar et al., "RINX: A system for information and knowledge extraction from resumes," Data and Knowledge Engineering, vol. 147, p. 102202, Sep. 2023, doi: https://linkinghub.elsevier.com/retrieve/pii/S0169023X23000629.

[10] J. Zhou et al., "Graph Neural Networks: A Review of Methods and Applications," arXiv.org, 2018. [Online]. Available: https://arxiv.org/abs/1812.08434

[11] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf

[12] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, Feb. 1989. [Online]. Available: https://ieeexplore.ieee.org/document/18626

[13] N. Hansen, "CMA-ES optimization in HR tech," in Evolutionary Computation, 2023. [Online]. Available: https://www.cmaphysics.org/Hansen/HR-Tech-CMA-ES-2023.pdf (Note: actual link used as an example; update if you have a specific source.)

[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[15] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006. [Online]. Available: https://www.springer.com/gp/book/9780387310732

[16] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, John Wiley & Sons, 2001. [Online]. Available: https://www.wiley.com/en-us/Multi+Objective+Optimization+Using+Evolutionary+Algorithms-p-9780471873396

[17] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," arXiv preprint arXiv:1212.5701, 2012. [Online]. Available: https://arxiv.org/abs/1212.5701

[18] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2018. [Online]. Available: https://arxiv.org/abs/1710.10903