



南开大学统计与数据科学学院 生物医学平台

目录

总体介绍.....	3
团队介绍.....	4
生物统计学团队	4
生物信息学团队	4
数学与流行病学团队	4
计算机视觉与影像分析团队	4
自然语言处理与大模型研究团队	4
生物统计学团队.....	5
邓路	5
付盛	7
韩东啸.....	8
武颖	10
张巧真.....	12
陈博	13
王景丽.....	15
生物信息学团队.....	16
胡刚	16
郑伟	18
王奎	20
宿鸿	21
游榕晖.....	23
赵尉辰.....	24
数学流行病学团队	25
黄森忠.....	25
计算机视觉与影像分析团队	26
潘斌	26
徐君	27

自然语言处理与大模型研究团队	28
李慧嘉.....	28
王冀淇.....	29
张瑶	30

总体介绍

南开大学统计学科是上世纪五十年代，在王梓坤院士和胡国定先生奠定的概率论与数理统计专业基础上发展起来的。南开统计拥有全国首批统计学一级学科博士学位授予权，2012 年第三轮学科评估中并列全国第四，并入选全国首批统计学博士后流动站。2017 年和 2022 年南开统计学科两次均入选国家“双一流”建设学科，在第四轮学科评估中并列全国第三（A）。

南开统计在生物医学相关方向的研究主要涉及生物统计、生物信息学、数学流行病学、计算机视觉和自然语言处理在生物医学中的应用等，近年来在生物医学权威期刊 Nature Biotechnology、Nature Cancer、Circulation、Gut、Nature Communications，生物统计和生物信息权威期刊 Journal of the American Statistical Association、Biometrika、Bioinformatics、Biometrics、Briefings in Bioinformatics、PLoS Computational Biology 以及人工智能权威期刊和会议 Journal of Machine Learning Research、IEEE TIT、IEEE TIP、ICML、NIPS、ACL 等发表数十篇论文，多篇疫情研判报告获得国家领导人批示，在国际蛋白质结构预测和功能预测竞赛中多年多次获得冠军，相关成果获得天津市自然科学一等奖和天津市科技进步二等奖。

南开统计与数据科学学院生物医学平台面向人民生命健康，依托传染病溯源预警与智能决策全国重点实验室、天津市医药数据分析与统计研究重点实验室，致力于将统计和人工智能方法深入应用到生物医学数据分析中。研究开发的方法广泛应用于生命科学、临床医学、药物研发、公共卫生等领域，与宾夕法尼亚大学佩雷尔曼医学院、中国医学科学院阜外医院、天津市第一中心医院、南开大学生命科学院、医学院等生物、医学领域专家合作发表多项重要成果。欢迎生物医学领域和医药企业领域的专家与我们深入合作，共同迎接生物医学大数据和人工智能时代的来临。

团队介绍

生物统计学团队目前包括 7 位 PI，研究涉及生存分析、全基因组关联数据分析、医学诊断和预测、精准医疗、临床试验、复杂生物医学数据分析等多个领域。近年来在统计、生物统计以及生物信息学权威期刊 *Journal of the American Statistical Association*、*Biometrika*、*Biometrics*、*Bioinformatics*、*PLOS Computational Biology* 以及机器学习和信息论权威期刊 *Journal of Machine Learning Research* 和 *IEEE TIT* 等发表论文数十篇，并与天津医科大学、天津中医药大学、南开大学第一附属医院（天津市人民医院）、天津市第一中心医院和天津市眼科医院等开展了广泛的合作研究，研究成果发表于 *Gut* 等临床权威期刊。

生物信息学团队目前包括 6 位 PI，研究涉及多组学数据分析、蛋白质结构与功能预测、蛋白质设计、图深度学习等领域。近年来在 *Nature* 系列子刊 *Nature Biotechnology*、*Nature Cancer*、*Nature Communications*，临床医学权威期刊 *Circulation*、*Acta Neuropathologica*，生物信息学权威期刊 *Bioinformatics*、*Briefings in Bioinformatics* 等发表论文数十篇。与南开大学生命科学院、南开大学医学院、国家疾控中心、宾夕法尼亚大学、中国医学科学院阜外医院、天津中医药大学、南开大学第一附属医院（天津市人民医院）、天津市第一中心医院和天津市眼科医院等开展了广泛的合作研究。团队多次获得国际蛋白质结构预测竞赛(CASP)、国际蛋白质固有无序区域预测竞赛(CAID)、国际蛋白质功能预测竞赛(CAFA)第一名，获得天津市自然科学一等奖。

数学与流行病学团队目前包括 1 位 PI 黄森忠教授团队，研究领域为数学流行病学以及泛函分析及应用。新冠疫情以来，团队向国务院、卫健委、国家疾控中心提交了数百份疫情研判报告，多封报告获得国家领导人批示。

计算机视觉与影像分析团队目前包括 2 位 PI，研究领域为图像处理、计算机视觉、高效图像修复，编辑和生成。近年来在计算机视觉和图像处理权威期刊 *IEEE TIP*、*IEEE TGRS*、*IEEE TMM*、*PR* 以及计算机顶会 *ICML*、*NeurIPS*、*CVPR*、*ICCV*、*ECCV* 等发表成果 50 篇以上，多篇论文入选 ESI 热点论文和高被引论文。研究成果应用在新冠肺炎 CT 筛查和天津城南变电站、贵州高速、烟台交管等部门的项目中。

自然语言处理与大模型研究团队目前包括 3 位 PI，研究领域为自然语言处理、数据挖掘以及领域知识增强等。近年来在 *IEEE TKDE*、*PNAS* 以及人工智能和自然语言处理顶会 *ACL*、*EMNLP*、*AAAI* 发表多篇论文，获得天津市科技进步二等奖。

生物统计学团队

邓路：副教授，硕士生导师，主要研究方向为生物统计学，包括孟德尔随机化分析在高维全基因组关联数据中的理论和应用，以及基于汇总统计量的数据整合方法的研究，例如多源数据的隐私保护与整合分析，医学诊断和预测等，已在 *Biometrika*, *Bioinformatics*, *Biometrics*, *Statistica Sinica* 等期刊杂志发表论文 10 余篇。

联系方式： denglu014@mail.nankai.edu.cn

研究案例 1：数据来自于产妇中心性肥胖对其婴儿出生体型和青春期身高增长的因果研究。我们考虑三个代表产妇中心性肥胖的风险因素，包括腰臀比、腰围和臀围。婴儿出生体型结果则用出生体重、出生身长和头围来表示。研究目标是通过逐个分析这 9 对风险因素-结果之间的因果效应，来探究产妇中心性肥胖与其婴儿出生体型之间的因果关联。

引入遗传基因变量作为工具变量，根据工具变量选取准则，我们选择了 41 个与风险因素腰臀比具有显著统计相关性的独立的单核苷酸多态性 (SNPs) 作为工具变量。数据来自两组 GWAS 研究。一组观测了婴儿出生体型结果和其对应的 41 个 SNPs；另一组观测了产妇的三个中心性肥胖风险因素和其对应的 41 个 SNPs。

研究案例 2：数据来自于人类睡眠行为的基因遗传基本的研究。我们采用由加速度计衍生的测量方法，量化了八种人类睡眠行为 (Diurnalinactivity, L5timing, M10timing, Num.nocturnalsleepepisodes, Sleepduration, Sleepdurationvariability, Sleep efficiency, Sleepmidpointtiming)。研究目标是通过逐个估计腰臀比 (Waist-hip-ratio) 与上述八种睡眠行为的因果效应，来评估腰臀比是否会对人类睡眠行为产生因果影响。

我们选择与腰臀比和睡眠行为有关联的一些单核苷酸多态性 (SNPs) 作为工具变量。SNPs 来自公开的 European reference genomes projects。使用 Linkage Disequilibrium 技术选取统计意义上独立的 SNPs 之后，最终获得了 95, 819 个 SNPs。

代表性成果：

1. **Lu Deng**, Sheng Fu, Jing Qin and Kai Yu. (2024) On combining individual-level data with summary data in statistical inferences. *Statistica Sinica*, 34, 1505-1520.
2. Sheng Fu, **Lu Deng**, Han Zhang, William Wheeler, Jing Qin and Kai Yu. (2023). Integrative analysis of individual-level data and high-dimensional summary statistics. *Bioinformatics*, 39(4): btad156.
3. **Lu Deng**, William Wheeler, Kai Yu. (2023). Mendelian randomization test of causal effect using high-dimensional summary data. *Statistica Sinica*, 33, 1365-1387.

4. Han Zhang, **Lu Deng**, Mark Schiffman, Jing Qin, Kai Yu. (2020). Generalized integration model for improved statistical inference by leveraging external summary data. *Biometrika*, 107, 689-703.
5. **Lu Deng**, Han Zhang, Lei Song and Kai Yu. (2020) Approximation of bias and mean squared errors in two-sample Mendelian randomization analyses. *Biometrics*, 76, 369-379.

付盛：副教授，硕士生导师，主要从事复杂数据的统计建模与应用研究。研究方向包括：医疗数据的精准分类与个性化医疗推荐，基于隐私保护的多源数据整合分析与知识迁移方法。相关成果发表于 *IEEE Transactions on Information Theory*、*Statistica Sinica*、*Bioinformatics*、*PLOS Genetics*、*PLOS Computational Biology* 等国际知名期刊，并曾获得美国国立卫生研究院（NIH）颁发的杰出研究奖。

在精准医疗方面，我与合作者提出了一系列分类与个体化推荐算法，能够结合患者的临床特征和诊疗记录，识别疾病亚型、制定最优治疗方案。例如，我们发展了鲁棒的个体化治疗规则学习方法，以及适应性加权的多分类模型，特别适用于真实医疗数据中存在噪声或类别不平衡的情况，为实现个性化诊疗提供了有力工具。

在隐私保护与多源数据整合方面，我们开发了多种基于汇总统计量的分析方法，无需访问原始个体数据，即可实现高效、稳健的跨机构数据整合。这类方法特别适合多中心合作或数据共享受限的情境，如疾控系统中的跨区域疫情监测与慢病管理。同时，我们也开展了统计迁移学习研究，旨在实现不同人群、不同数据来源之间的模型迁移与共享，提升分析方法的泛化能力和适应性。

联系方式： fusheng@nankai.edu.cn

代表性论文：

1. **Fu, S.**, Wheeler, W., Wang, X., Hua, X., Godbole, D., Duan, J., ... & Yu, K. (2024). A comprehensive framework for trans-ancestry pathway analysis using GWAS summary data from diverse populations. *PLoS genetics*, 20(10), e1011322.
2. **Fu, S.**, Purdue, M. P., Zhang, H., Qin, J., Song, L., Berndt, S. I., & Yu, K. (2023). Improve the model of disease subtype heterogeneity by leveraging external summary data. *PLOS Computational Biology*, 19(7), e1011236.
3. **Fu, S.**, Deng, L., Zhang, H., Wheeler, W., Qin, J., & Yu, K. (2023). Integrative analysis of individual-level data and high-dimensional summary statistics. *Bioinformatics*, 39(4), btad156.
4. **Fu, S.**, He, Q., Zhang, S., & Liu, Y. (2019). Robust outcome weighted learning for optimal individualized treatment rules. *Journal of biopharmaceutical statistics*, 29(4), 606-624.
5. **Fu, S.**, Chen, P., & Ye, Z. (2023). Simplex-based proximal multicategory support vector machine. *IEEE Transactions on Information Theory*, 69(4), 2427-2451.

韩东啸：南开大学统计学院副教授，主要研究方向为生存分析、高维数据统计推断、机器学习理论研究等。在统计学顶级期刊 *Journal of the American Statistical Association*、机器学习顶级期刊 *Journal of Machine Learning Research*、计量经济学顶级期刊 *Journal of Econometrics*、经济统计顶级期刊 *Journal of Business and Econometric Statistics*、生物统计顶级期刊 *Biometrics* 等发表十余篇论文。其中发表在 *Biometrics* 上的论文被评为 *Biometrics* 期刊 2020-2021 年度高被引用论文。主持国家自然科学基金面上项目、青年项目各一项，主持天津市面上项目一项。曾指导学生获得北美数学建模比赛 F 奖。现为中国现场统计研究会贝叶斯分会常务理事、全国工业统计学教学研究会数字经济与区块链分会常务理事。

联系方式：handongxiao@nankai.edu.cn

主要创新成果 1：生存数据的统计分析。我们关于带有连续标记的生存数据建立了可加风险模型，并给出相应的参数估计、假设检验和变量选择方法。这些成果分别发表在统计学重要期刊 *Statistical Methods in Medical Research* (2021) 和统计学期刊 *Lifetime Data Analysis* (2017) 上。这些方法可用于分析 HIV 疫苗效力数据、流感病毒疫苗效力数据等常见的临床数据，对于疫苗的研发有着重要的意义和指导作用。此外，我们针对单纯删失数据和伴有复发事件的数据建立了带有随机效应的多种生存模型，并给出了参数估计和变量选择的方法。相关工作分别发表在生物统计顶级期刊 *Biometrics* (2020)、医学统计学重要期刊 *Statistical Methods in Medical Research* (2019, 2021) 以及 *Statistics in Medicine* (2021) 上。这些研究为癌症复发的诱因研究以及基因对疾病影响的探索提供了重要的理论基础和技术支持，其成果在医疗研究领域具有重要的指导意义。

主要创新成果 2：大数据的稳健的统计分析。我们基于 Huber Loss 损失函数，分别给出了高维线性模型、高维单指标模型、带有成分协变量 (compositional covariates) 的高维线性对数对比模型 (high-dimensional linear log-contrast models) 的支撑集恢复 (support recovery)、逐点的统计推断方法以及群组的统计推断方法。这些工作分别发表在计量经济顶级期刊 *Journal of Econometrics* (2022)、经济统计顶级期刊 *Journal of Business and Econometric Statistics* (2023) 以及统计学期刊 *Scandinavian Journal of Statistics* (2023) 上。另外，我们还在流数据的框架下，基于一般的损失函数给出了高维回归参数的逐点的统计推断方法，该方法包含 Huber Loss 损失函数作为特例。这个研究成果发表在 *Journal of Machine Learning Research* (2024) 上。我们的这些研究显著提升了模型的可解释性和稳健性，为实际经济数据、医学数据的处理提供了重要的理论指导与实践参考，对于解决经济领域中的复杂问题具有深远意义。

代表性论文:

1. **Han,D.**, Zheng, S., Shen, G., Song, X., Sun, L., and Huang, J. (2025). Deep Mutual Density Ratio Estimation with Bregman Divergence and Its Applications. *Journal of the American Statistical Association*. accepted.
2. **Han, D.**, Xie, J., Liu, J., Sun, L., Huang, J., Jian, B., and Kong, L. (2024). Inference on high-dimensional single-index models with streaming data. *Journal of Machine Learning Research*, 25, 1-68.
3. **Han,D.**, Huang, J, Lin, Y., Liu, l.,Qu, L., and Sun, L. (2023). Robust Signal Recovery for High-dimensional Linear Log-contrast Models with Compositional Covariates. *Journal of Business and Econometric Statistics*, 41, 957-967.
4. **Han, D.**, Huang, J., Lin, Y., and Shen, G. (2022). Robust post-selection inference of high-dimensional mean regression with heavy-tailed asymmetric or heteroskedastic errors. *Journal of Econometrics*, 230, 416-431.
5. **Han, D.**, Su, X., Sun, L., Zhang, Z., and Liu, L. (2020). Variable selection in joint frailty models of recurrent and terminal events. *Biometrics*, 76, 1330-1339.

武颖：副教授，硕士生导师，主要研究方向为生物医药数据的统计分析、生存分析、纵向数据分析、临床试验等。

联系方式：yw@nankai.edu.cn

研究案例 1：与天津市医科大学和天津市第一中心医院等团队合作研究关于肝细胞癌复发生存临床预测统计模型的合作项目中，通过使用疾病死亡模型等多状态模型，识别在肝细胞癌治疗后影响其预后效果的不同治疗方案（肝移植、肝切除、射频治疗）、临床相关参数、肝功能指标、肿瘤状态及遗传因素。该研究可用来探索影响肝细胞癌复发和死亡的危险因素，预测病人在治疗后 2 年、3 年、5 年复发率和死亡率，有望在临床层面上辅助优化早期肝细胞癌个体化多学科治疗决策，该成果发表于临床医学权威期刊 Gut。

研究案例 2：中药注射剂是传统中医药理论与现代生产工艺相结合的产物，突破了中药传统的给药方式，是中药现代化的产物。然而，中药成分复杂，随着中药注射剂应用日渐广泛，其引起的药物不良反应也逐渐增多。安全性和有效性是药品的两个基本属性，药品上市后的再评价对了解不良反应率及其人群特征和影响因素等具有重要意义，为医生和患者安全用药提供科学依据。关于药物不良反应的分类与预测，其中一个主要困难是数据极端不平衡，阳性样本与阴性样本的比率大约在 1/1000。我们基于中医药电子病历大数据，关于中药注射剂再上市临床安全性的统计及机器学习方法分析，首先使用自然语言处理等方法将非结构化文本大数据转化成可供分享的结构化数据，之后采用多种机器学习和深度学习方法预测极端不平衡事件。该部分成果发表在 2021 ICMHI 会议期刊上。

代表性成果：

1. Zhang, N.*, Jiang, W. *, Zhang, Y. *, Song, T.Q. *, Lv, J. *, Gu, J. *, **Wu, Y. ***, Qian, J. *, Tian, D., Guo, Q., Zhang, L. et al. (2020) Individualised tailored assessment of therapeutic alternatives for HCC patients within the Milan criteria. *Gut*, 69(10), 1893-1895. (* co-first author)
2. Zhao, Y. X., Yuan, H., and **Wu, Y.** (2021). Prediction of adverse drug reaction using machine learning and deep learning based on an imbalanced electronic medical records dataset. In *2021 5th International Conference on Medical and Health Informatics*, pages 17 - 21.
3. Lv, J., Zhang, N., Du, Y., **Wu, Y.**, Song, T.-Q., Zhang, Y., Qu, Y., Liu, Y., Gu, J., Wang, Z., Qiu, Y., Yang, B., Tian, D., Guo, Q., Zhang, L., Sun, J., Xie, Y., Wang, Z., Sun, X., Jiang,

- W., and Lu, W. (2021). Comparison of liver transplantation and liver resection for hepatocellular carcinoma patients with portal vein tumor thrombus type I and type II. *Yonsei Medical Journal*, 62(1):29 – 40.
4. He, J., Hu, S., Xu, X., Guo, P., Niu, Y., Zhang, J., Zhang, R., Chen, S., Ma, S., Liu, F., Li, Q., Li, C., Zhang, L., **Wu, Y.**, Zhang, M., Zhang, M. (2022). Association of long-term exposure to PM2. 5 in workplace with fasting plasma glucose among asymptomatic adults: A multicenter study in North China. *Environment International*, 166, 107353.
 5. Iraniparast, M., Shi, Y., **Wu, Y.**, Zeng, L., Maxwell, C. J., Kryscio, R. J., John, P. D. S., SantaCruz, K. S., and Tyas, S. L. (2022). Cognitive reserve and mild cognitive impairment: Predictors and rates of reversion to intact cognition vs progression to dementia. *Neurology*, 98(11), e1114-e1123.。

张巧真：南开大学统计与数据科学学院副教授，试验设计团队成员。主要研究方向为试验设计，具体包括超饱和设计、计算机试验设计、序贯试验设计等，已在 *Proceedings of Machine Learning Research*、*Journal of Quality Technology*、*Journal of Multivariate Analysis* 等期刊发表高水平学术论文 10 余篇。已主持国家自然科学基金 1 项、国家重点实验室开放课题 1 项。现任中国数学会均匀设计分会理事，中国现场统计研究会试验设计分会理事，天津市现场统计研究会常务理事等。

联系方式：zhangqz@nankai.edu.cn

代表性成果：

1. Liu, S., Cao, Y., **Zhang, Q.**, Feng, L., An, B., Mitigating Underfitting in Learning to Defer with Consistent Losses, *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024*, Valencia, Spain. PMLR: Volume 238.
2. Liu, S., **Zhang, Q.**, Cao, Y., Feng, L., An, B. Consistent Complementary-Label Learning via Order-Preserving Losses. *Proceedings of Machine Learning Research* 2023, 206, 8734-8748.
3. Zheng C. W., Qi Z. F., **Zhang Q. Z.**, Liu M. Q. A Method for Augmenting Supersaturated Designs with Newly Added Factors. *Mathematics* 2023, (2227-7390) 11(1):60.
4. Huang, H. Z., Lin, D. K. J., Liu, M. Q. and **Zhang, Q. Z.** Variable selection for kriging in computer experiments. *Journal of Quality Technology* 2020, 52(1), 40-53.
5. **Zhang, Q. Z.**, Dai, H. S., Liu, M. Q. and Wang, Y. A method for augmenting supersaturated designs. *Journal of Statistical Planning and Inference* 2019,199, 207-218.

陈博：特聘副研究员，多伦多大学统计系本科，硕士，博士。2018-2022 年在加拿大玛格丽特公主癌症中心从事博士后研究，2023 年起南开大学工作。主要研究复杂生物统计数据的分析方法，在 *Biostatistics*, *PLOS Computational Biology* 等杂志共发表文章 10 篇，主持国家自然科学基金青年基金一项，对生物统计领域的复杂数据的统计方法进行了大量创新性研究，解决了数据本身维度高，相关性强，不服从传统统计分布及多次重复观测等问题。在全基因组关联性分析的问题上，X 染色体上基因的关联性分析方法研究工作因为问题复杂，理论难度大，曾长期处于几乎停滞状态。我们提出的两项开创性的方法，解决了这一重要的长期公开问题，使这一研究领域重新焕发活力。这两项工作受到业内广泛关注，其中一项工作被 *Genetic Epidemiology* 期刊统计为“被引用最多的文章”。在微生物组测序数据分析问题上，我们同样提出了多项革新性的方法并被广泛应用，对微生物组测序数据的分析方法做出了突破性的贡献。

联系方式：bochen@nankai.edu.cn

与生物医学合作研究案例：我参与了加拿大玛格丽特公主癌症中心的许多临床问题研究，为临床研究提供了统计方法上的支持。代表性案例如下：

研究案例 1：微生物生态系统疗法与免疫检查点抑制剂联合治疗晚期实体瘤患者的疗效分析。在传统的免疫疗法（免疫检查点抑制剂 ICI）的基础上，我们首次提出了一种联合 ICI 的全新的微生物生态系统疗法（MET4），针对样本量小但维度高的微生物组数据特点，分析了新疗法和传统 ICI 疗法相比人体肠道微生物组组成的区别，证明了微生物菌群作为癌症 ICI 治疗的辅助治疗手段的进一步发展是合理的。相关成果发表在肿瘤学顶刊 *Annals of Oncology*。

研究案例 2：肺移植接受者的微生物组与胃食管反流、炎症和肺移植功能障碍的关系的研究。胃食管反流病（GERD）如何影响肺移植受体的移植器官微生物群落组成及其对肺移植验证和功能的影响尚不清楚。我们比较了移植后第一年内肺移植受体的移植器官微生物群落，并评估 GERD、移植器官微生物群落，炎症和肺移植功能障碍之间的关联。运用对纵向微生物组数据的统计分析方法，找到了受 GERD 密切影响的菌群，并分析出哪些菌群会导致炎症和肺移植功能障碍的早期发展。这一成果发表于美国呼吸与重症监护医学权威杂志 *AJRCCM (The Blue Journal)*。

代表性成果：

1. Anna Spreafico; Alya A. Heirali; Daniel V. Araujo; Tira J. Tan; Marc Oliva; Pierre H. H. Schneeberger; **Bo Chen**; Matthew K. Wong; Lee-Anne Stayner; Aaron R. Hansen; Samuel D. Saibil; Ben X. Wang; Kyla Cochrane; Keith Sherriff; Emma Allen-Vercoe; Wei Xu;

- Lillian L. Siu; Bryan Coburn; First-in-class Microbial Ecosystem Therapeutic 4 (MET4) in combination with immune checkpoint inhibitors in patients with advanced solid tumors (MET4-IO Trial), *Annals of Oncology*, 2023, 34(6): 520-530.
2. Pierre H. H. Schneeberger; Chen Yang Kevin Zhang; Jessica Santilli; **Bo Chen**; Wei Xu; Youngho Lee; Zonelle Wijesinha; Elaine Reguera-Núñez; Noelle Yee; Musawir Ahmed; Kristen Boonstra; Rayoun Ramendra; Courtney W. Frankel; Scott M Palmer; Jamie L. Todd; Tereza Martinu; Bryan Coburn; Lung Allograft Microbiome Association with Gastroesophageal Reflux, Inflammation, and Allograft Dysfunction, *American Journal of Respiratory and Critical Care Medicine*, 2022, 206(12): 1495-1507.
 3. **Bo Chen**; Radu V. Craiu; Lisa J. Strug; Lei Sun; The X factor: A robust and powerful approach to X-chromosome-inclusive whole-genome association studies, *Genetic Epidemiology*, 2021, 45(7): 694-709.
 4. **Bo Chen**; Radu V Craiu; Lei Sun; Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study, *Biostatistics*, 2020, 21(2): 319-335.
 5. **Bo Chen**; Wei Xu; Generalized estimating equation modeling on correlated microbiome sequencing data with longitudinal measures, *PLOS Computational Biology*, 2020, 16(9): e1008108.

王景丽：讲师，硕士生导师，主要研究方向为精准医学、亚组识别以及孟德尔随机化分析等。主持国家自然科学基金青年项目。相关成果发表在生物统计权威期刊 *Statistics in Medicine* 等。

联系方式：jllwang@nankai.edu.cn

与医学相关的案例：在精准医学的领域中，我们提出了一种基于线性模型的多阈值变点决策方法，用于识别患者亚组，特别适用于对不同治疗方案反应不同的群体。该方法在系统性硬化症临床试验中成功区分了口服牛胶原蛋白治疗反应不同的患者，并通过临床指标分析验证了治疗效果差异。在乳腺癌研究中，该方法分析了 24,000 多个基因表达数据，识别出与肿瘤大小或转移相关的患者亚组。这一方法可推广至神经科学、心血管疾病等领域，为个性化医疗提供有力支持。

为了解决模型内生性问题（即未观测变量对目标变量的潜在影响），我们提出了一种基于变点分析的分段结构方程模型，能够自动将研究对象划分为不同的亚组进行分析。该方法在菲律宾家庭食物消费研究中得到验证：当母亲营养知识高的家庭收入对食物消费的影响强度是低知识水平家庭的两倍，揭示了营养知识在调节收入-食物消费关系中的关键作用。这一创新方法具有广泛的适用性。在医学随机对照试验中，它能有效处理患者不依从治疗方案导致的内生性问题，提高治疗效果评估的准确性。在遗传流行病学领域，该方法可增强孟德尔随机化研究的效力，通过亚组划分更精确地解析基因变异与疾病的因果关系。

代表性成果：

1. **Jingli Wang**, Jialiang Li, Yaguang Li & Weng Kee Wong. A model-based multi-threshold method for subgroup identification, *Statistics in Medicine*, 2019,38(14):2605-2631.
2. **Jingli Wang** & Jialiang Li. Multi-Threshold Structural Equation Model, *Journal of Business & Economic Statistics*, 2023, 41(2):377-387.

生物信息学团队

胡刚：教授，博士生导师，数据科学系主任，传染病溯源预警与智能决策全国重点实验室 PI，主要研究方向为生物信息学，包括组学数据分析、蛋白质结构和功能预测以及蛋白质设计等，获得天津市自然科学一等奖一项，主持国家自然科学基金重大研究计划培育项目、天元基金交叉重点专项、面上项目等项目多项。以第一作者或者通讯作者(含并列)在 *Nature Cancer*, *Circulation*, *Nature Communications*, *Bioinformatics* 等杂志发表学术论文多篇。开发的单细胞组学数据去噪分析工具 DESC 被广泛引用，入选 ESI 高被引论文。开发的无序区域预测的 flDPnn 系列算法获得首届和第二届国际蛋白质无序区域预测竞赛第一名，flDPnn 的论文被 *Nature Communications* 编辑列为 Featured articles，同样入选 ESI 高被引论文，同时入选国际著名学术组织“Faculty Opinions”的推荐论文。

联系方式：huggs@nankai.edu.cn

与生物医学合作研究案例：我们与中国医学科学院阜外医院利用单细胞转录组测序技术首次描绘了不同阶段自身免疫性心肌炎小鼠模型心脏免疫微环境特征，构建了不同阶段自身免疫心肌炎动物模型模拟心肌炎从发病到转变为心肌病的整个过程，并利用自己开发的单细胞转录组分析方法 DESC 等对这些单细胞数据进行了深入的分析，发现了心肌炎炎症阶段相关的细胞类型均高表达转录因子-Hif1a，此后我们对不同阶段自身免疫心肌炎小鼠给予 Hif1a 抑制剂（PX-478）进行治疗，发现该靶向抑制剂可以有效地抑制小鼠心脏炎症反应，该研究为临床治疗心肌炎提供了重要的潜在治疗策略。此项工作的前期结果在医学权威杂志 *Circulation* 发表；

我们与宾夕法尼亚大学合作，通过单核转录组研究了老年痴呆症的免疫机理，揭示了小胶质细胞在老年痴呆症中的作用，成果发表于神经科学顶刊 *Acta Neuropathologica*；另外，我们与南开生科院合作关于结直肠癌耐药性方面的研究发表于 *Nature Cancer*。

代表性成果：

1. Zheng H, Liu J, Cheng Q, Zhang Q, Zhang Y, Jiang L, Huang Y, Li W, Zhao Y, Chen G, Yu F, Liu L, Li Y, Liao X, Xu L, Xiao Y, Zheng Z, Li M, Wang H, **Hu G***, Du L*, Chen Q*. Targeted activation of ferroptosis in colorectal cancer via LGR4 targeting overcomes acquired drug resistance. *Nature Cancer*. 2024 Jan 30.
2. Zhao Y, Wang K, **Hu G***. DIST: spatial transcriptomics enhancement using deep learning. *Briefings in Bioinformatics*. 2023 Mar 19;24(2):bbad013.
3. **Hu G#**, Katuwawala A#, Wang K#, Wu Z, Ghadermarzi S, Gao J, Kurgan L. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions.

Nature Communications. 2021 Jul 21;12(1):4438. **ESI Highly Cited Paper. Featured among the Editor's Highlights. Faculty Opinions Recommended.**

4. Hua X#, **Hu G#**, Hu Q#, Chang Y, Hu Y, Gao L, Chen X, Yang PC, Zhang Y*, Li M*, Song J*. Single-Cell RNA Sequencing to Dissect the Immunological Network of Autoimmune Myocarditis. *Circulation*. 2020 Jul 28;142(4):384-400.
5. Li X, Wang K, Lyu Y, Pan H, Zhang J, Stambolian D, Susztak K, Reilly MP, **Hu G***, Li M*. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nature Communications*. 2020 May 11;11(1):2338. **ESI Highly Cited Paper.**

郑伟：教授，博士生导师，国家级海外青年人才，南开大学百名青年学科带头人，传染病溯源预警与智能决策全国重点实验室成员。蛋白质预测结构文件存储格式 ModelCIF 的国际标准制定委员会委员。

联系方式：jlspzw@nankai.edu.cn

学术介绍：

在密歇根大学从事博士后研究，研究方向为基于新型人工智能的全原子分子互作结构预测算法，主持开发的 C-I-TASSER、C-QUARK、D-I-TASSER、DMFold 等蛋白质单体结构预测算法、蛋白质-蛋白质互作复合物结构预测算法、核酸-核酸互作复合物结构预测算法、蛋白质-核酸复合物结构预测算法、生物分子多构象预测算法等，常年、多次获得被誉为“蛋白质结构预测领域的奥林匹克竞赛”的国际赛事（CASP）的多项冠军，领先包括 AlphaFold2、AlphaFold3 在内的全球 80 余个学术界及工业界的课题组。多次受邀在世界蛋白质结构预测大赛赛后国际会议、世界生物医药与人工智能大会做特邀报告。累计在 Nature Biotechnology、Nature Methods、Nature Communications、Nature Protocols、Science Advances、Nature Computational Science、Nucleic Acids Research、PNAS 等高水平 SCI 期刊发表文章 50 余篇。据 Google Scholar 记录，相关研究成果已累计获得 3400 余次引用。担任 SCI 期刊 Molecules 杂志特约编辑及 Nature Communications、Nature Machine Intelligence、Nature Computational Science 等 SCI 期刊审稿人。主导开发的算法已经累计服务了超过 100 个国家的 10 万多名用户，发布了 SARS-CoV-2 病毒的全蛋白质组预测结构模型，受到了《纽约时报》等百余个杂志的报道，已有 100 万阅读量。DMFold 算法在 CASP15 的抗体-抗原复合物预测上达到了媲美实验解析的精度，Nature 对此做了特别报道。DMFold 算法被 BioArt、DrugAI 等数十个生物学公众号转载，目前累计阅读量约 3 万。主持并参与了多个国家级、天津市级人才项目及重点项目。

研究方向：

- （1）蛋白质等生物大分子结构、功能、互作预测研究；
- （2）抗体类药物筛选与优化研究，抗肿瘤多特异抗体计算筛选优化工作；
- （3）罕见病致病基因预测；
- （4）X 病毒预测、病毒感染性预测；
- （5）对病毒疫苗及中和抗体药物计算筛选、优化等。

研究实例：

示例 1：抗肿瘤多特异抗体、抗病毒中和抗体等计算筛选临床前研发

与某生物制药公司合作，对肿瘤的关键性蛋白 CD3、CD28、NKP46 等蛋白质筛选特异性结合抗体，进而研发对肿瘤多特异性抗体。上用上述抗原蛋白免疫动物、在动物体内提取测序抗体序列，数量级在 20-30 万左右。首先使用基于深度学习及高维统计的工具将数据样本聚类，获得 2-3 万非冗余的蛋白质序列。使用自研的世界领先的蛋白质复合物结构预测算法，预测每一个抗体序列与抗原蛋白序列形成的复合物的三维空间结构。再根据置信系数对 2 万个抗体序列结合抗原的可能性进行排序。下一步根据预测的置信系数及预测结构的表位信息，选取排名较高的约 200 组抗体进行 ELISA 等生物实验验证。最终获得 30 余株有效的结合抗体。

示例 2：结合冷冻电镜数据进行生物分子实验结构解析

与南开大学药物化学生物学全国重点实验室合作，针对国内目前最高生物防护等级的四级病原—克里米亚-刚果出血热（新疆出血热）病毒的转录复制机制开展研究，应用自行开发的 DMFold 等结构预测算法，在不依赖实验数据的情况下，获得了该病毒核蛋白-核酸复合体这一超大型结构的模型，经与冷冻电镜解析获得的天然复合体结构进行比较与迭代，解析了超高分辨率的蛋白-核酸复合体结构。

代表性论文：

1. **Wei Zheng**, Qiqige Wuyun, Yang Li, Quancheng Liu, Xiaogen Zhou, Chunxiang Peng, Yiheng Zhu, Lydia Freddolino, Yang Zhang (2025) Single- and multi-domain protein structure prediction with D-I-TASSER. *Nature Biotechnology*, doi:10.1038/s41587-025-02654-4. 5-years **IF=57.0** (第一作者)
2. **Wei Zheng**, Qiqige Wuyun, Yang Li, Chengxin Zhang, Lydia Freddolino, Yang Zhang. (2024) Improving deep learning protein monomer and complex structure prediction using DeepMSA2 with huge metagenomics data. *Nature Methods* 21: 279–289. 5-years **IF=45.6** (第一作者)
3. SM Mortuza#, **Wei Zheng**#, Chengxin Zhang, Yang Li, Robin Pearce, Yang Zhang (2021) Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. *Nature Communications* 12 (1), 5011. 5-years **IF=16.1** (共同一作)
4. Pengshuo Yang#, **Wei Zheng**#, Kang Ning, Yang Zhang. (2021) Decoding the link of microbiome niches with homologous sequences enables accurately targeted protein structure prediction. *Proceedings of the National Academy of Sciences* 118: e2110828118. 5-years **IF=10.8** (共同一作)
5. **Wei Zheng**, Qiqige Wuyun, Xiaogen Zhou, Yang Li, Lydia Freddolino, Yang Zhang. (2022) LOMETS3: integrating deep learning and profile alignment for advanced protein template recognition and function annotation. *Nucleic Acids Research* 50: W454-W464. 5-years **IF=16.1** (第一作者)

王奎：副教授，硕士生导师。主要研究方向为蛋白质结构分析、单细胞转录组数据与空间转录组数据分析及深度学习算法在生物数据中的应用。以第一作者或者通讯作者(含并列)在 *Acta Neuropathol*, *Nature Communications*, *Bioinformatics* 等杂志发表学术论文多篇。

联系方式： wangkui@nankai.edu.cn

学术介绍：蛋白质结构分析主要聚焦于蛋白质无序区域的预测问题。AI 模型在蛋白质结构预测中取得很大的成功，但是 AlphaFold 系列目前在对于无序区域的精度仍然不够。我们开发的 fIDPnn 系列算法历届国际蛋白质无序区域预测竞赛均名列前茅（两次第一名），是目前精度最高并且速度最快的预测算法。成果发表于 *Nature Communications* 和 *Journal of Molecular Biology*。

在组学数据分析中，研究兴趣主要集中于单细胞/空间转录组数据的聚类算法与数据整合方法的研究。同时对医学生物相关的单细胞/空间转录组数据的分析具有丰富的经验，我们与宾夕法尼亚大学合作，通过单核转录组研究了老年痴呆症的免疫机理，揭示了小胶质细胞在老年痴呆症中的作用，成果发表于神经科学顶刊 *Acta Neuropathologica*。

代表性论文：

1. **Wang K**, Hu G, Basu S, Kurgan L. fIDPnn2: Accurate and Fast Predictor of Intrinsic Disorder in Proteins. *Journal of Molecular Biology*. 2024 Sep 1;436(17):168605.
2. Hu G#, Katuwawala A#, **Wang K#**, Wu Z, Ghadermarzi S, Gao J, Kurgan L. fIDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications*. 2021 Jul 21;12(1):4438. **ESI Highly Cited Paper. Featured among the Editor's Highlights. Faculty Opinions Recommended.**
3. Nguyen AT#, **Wang K#**, Hu G#, Wang X, Miao Z, Azevedo JA, Suh E, Van Deerlin VM, Choi D, Roeder K, Li M, Lee EB. APOE and TREM2 regulate amyloid-responsive microglia in Alzheimer's disease. *Acta Neuropathologica*. 2020 Oct;140(4):477-493.
4. **Wang K**, Hu G, Wu Z, Su H, Yang J, Kurgan L. Comprehensive Survey and Comparative Assessment of RNA-Binding Residue Predictions with Analysis by RNA Type. *International Journal of Molecular Sciences*. 2020 Sep 19;21(18):6879.
5. **Wang K**, Hu G, Wu Z, Uversky VN, Kurgan L. Assessment of Disordered Linker Predictions in the CAID2 Experiment. *Biomolecules*. 2024 Feb 28;14(3):287.

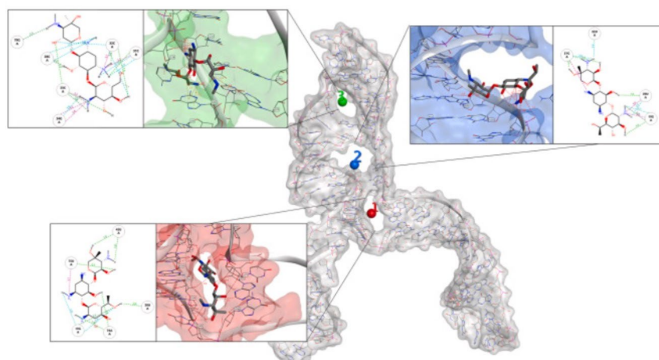
宿鸿：特聘副研究员，主要研究方向为生物信息学，包括蛋白质结构预测、蛋白质/RNA-配体结合位点预测和蛋白质功能网络预测等，主持德国洪堡基金博士后项目（洪堡学者），获得天津市优秀博士学位论文、南开大学优秀博士学位论文、南开大学优秀博士毕业生、全国博士后人工智能与发展论坛二等奖等荣誉。以第一作者或者共同第一作者在 Nature Protocols, Advanced Science, Bioinformatics 等杂志发表学术论文多篇。开发的蛋白质单体结构预测算法 trRosettaX，在国际蛋白质结构预测竞赛 CASP 中多次名列前茅（CASP15 夺得榜首），入选 ESI 高被引论文和热点论文。论文谷歌学术总引用 815 次。

联系方式：hong.su@nankai.edu.cn

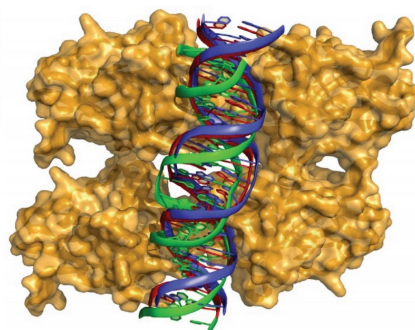
应用案例 1：图 A 展示了 RNA-配体结合算法预测算法 **RNA site** 在新冠病毒靶点识别中的实际应用，该案例源自洛桑大学医学院 Valeria Cagno 研究团队在 Antiviral Research 期刊发表的研究。图中，灰色表示新冠病毒的 RNA 结构，标记的三个点为 RNA site 预测的潜在药物结合位点。研究表明，这些结合位点正是小分子抗生素的作用位置，且基于 RNA site 的预测，研究人员找到了结合力更强的候选药物。

应用案例 2：图 B 展示了蛋白质-核酸结合位点预测算法 **NucBind** 在蛋白质-DNA 分子对接中的优化。图中，橙色表面表示肿瘤抑制因子 p53 的蛋白质结构，我们利用 NucBind 预测的 DNA 结合位点用于分子对接后（蓝色），所得结果与 DNA 分子参考结构（红色）非常接近，对接误差 RMSD 仅为 2.32 Å（该值越小，对接结果越好），而默认对接结果（绿色）则与参考结构存在较大偏差，RMSD 达 41.19 Å。

A. RNA site 在新冠病毒中的应用



B. NucBind 在分子对接中的应用



代表性成果:

1. **Su, H.#**, Wang, W.#, Du, Z., Peng, Z., Gao, S. H., Cheng, M. M., & Yang, J*. (2021). Improved protein structure prediction using a new multi-scale network and homologous templates. *Advanced Science*, 8(24), 2102592. (trRosettaX)
2. Du, Z.#, **Su, H.#**, Wang, W., Ye, L., Wei, H., Peng, Z., ... & Yang, J*. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nature protocols*, 16(12), 5634-5651. (trRosettaX server; ESI Highly Cited Paper.)
3. Zhu, K.#, **Su, H.#**, Peng, Z., & Yang, J*. (2023). A unified approach to protein domain parsing with inter-residue distance matrix. *Bioinformatics*, 39(2), btad070. (UniDoc)
4. **Su, H.**, Liu, M., Sun, S., Peng, Z., & Yang, J*. (2019). Improving the prediction of protein–nucleic acids binding residues via multiple sequence profiles and the consensus of complementary methods. *Bioinformatics*, 35(6), 930-936. (NucBind)
5. **Su, H.**, Peng, Z., & Yang, J*. (2021). Recognition of small molecule–RNA binding sites using RNA sequence and structure. *Bioinformatics*, 37(1), 36-42. (RNAsite)

游榕晖：特聘副研究员，主要研究方向为文本挖掘和生物信息学，包括蛋白质功能预测、超大规模多标签文本分类以及 MHC 分子绑定强度等。相关成果以第一作者身份发表在 NeurIPS、AAAI、ISMB、Nucleic Acids Research 以及 Bioinformatics 等机器学习和生物信息学领域的顶级会议和期刊上。开发的蛋白质功能预测算法 GOLabeleer 以及 NetGO 系列在功能预测的国际比赛 CAFA 中多次获得第一。同时，还开发了基于多标签注意力机制和低且宽的概率标签树的深度模型 AttentionXML。AttentionXML 能够处理百万级别样本和标签的超大规模数据，并且是首个性能超越传统方法的深度学习模型。相关成果以第一作者身份于 2019 年发表在机器学习领域的顶级会议 NeurIPS，并且根据 Google Scholar 记录被引次数达到 430 次。

联系方式：yourh@nankai.edu.cn

代表性成果：

1. **You R**, Yao S, Mamitsuka H, et al. DeepGraphGO: graph neural network for large-scale, multispecies protein function prediction[J]. *Bioinformatics*, 2021, 37(Supplement_1): i262-i271.
2. **You R**, Yao S, Xiong Y, et al. NetGO: improving large-scale protein function prediction with massive network information[J]. *Nucleic Acids Research*, 2019, 47(W1): W379-W387.
3. **You R**, Zhang Z, Wang Z, et al. AttentionXML: Label Tree-based Attention-aware Deep Model for High-performance Extreme Multi-label Text Classification[C]. *Advances in Neural Information Processing Systems*, 2019, 32.
4. **You R**, Liu Y, Mamitsuka H, et al. BERTMeSH: deep contextual representation learning for large-scale high-performance MeSH indexing with full text[J]. *Bioinformatics*, 2021, 37(5): 684-692.
5. **You R**, Qu W, Mamitsuka H, et al. DeepMHCII: a novel binding core-aware deep interaction model for accurate MHC-II peptide binding affinity prediction[J]. *Bioinformatics*, 2022, 38(Supplement_1): i220-i228.

赵尉辰：讲师，主要研究方向为图深度学习，生成模型，AI for Science，重点研究具有离散结构数据的表示学习、离散生成与离散采样算法、以及在自然科学中的应用，包括蛋白质结构和功能预测、蛋白质设计、多序列比对、药物分子设计等，主持国家自然科学基金青年基金一项，参与国家自然科学基金重大项目、国家重点研发计划、国家自然科学基金重点项目、天元基金交叉重点专项等项目多项。以第一作者在 KDD, Neural Networks 等人工智能顶会顶刊发表学术论文多篇。

联系方式： zhaoweichen@nankai.edu.cn

与生物医学合作研究案例： 作为骨干成员参与和中国科学院天津工业生物技术研究所合作的国家自然科学基金天元基金交叉重点专项项目——指导蛋白质设计与改造的 AI 模型与系统。负责蛋白质改造算法中图神经网络的设计与理论分析。

作为骨干成员参与和中国科学院深圳先进技术研究院、人民大学合作的国家自然科学基金天元基金交叉重点专项项目——结合高维统计与人工智能的抗体类药物筛选、优化及脱靶研究。负责抗原表位、抗体决定簇预测算法中图 Transformer 模型的设计、以及抗体-抗原复合物结构预测算法中生成扩散模型的设计与分析。

代表性成果：

1. **Zhao, W.,** Guo, T., Yu, X., & Han, C. (2023). A Learnable Sampling Method for Scalable Graph Neural Networks. *Neural Networks*, 162, 412-424.
2. **Zhao, W.,** Wang, C., Wang, X., Han, C., Guo, T., & Yu, T. (2025). Understanding Oversmoothing in Diffusion-Based GNNs From the Perspective of Operator Semigroup Theory. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '25)*. Association for Computing Machinery, New York, NY, USA, 2043 – 2054.
3. **Zhao, W.,** Wang, C., Han, C., & Guo, T. (2023). Exploring Over-smoothing in Graph Attention Networks from the Markov Chain Perspective. In *Proceedings of the 2023 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML '23)*. Association for Computing Machinery, New York, NY, USA, 282–288.

数学流行病学团队

黄森忠：教授，研究领域为数学流行病学，1998 年至 2018 年 3 月，在德国罗斯托克大学从事教学和科研。2018 年 4 月至 2020 年 6 月，南开大学统计与数据科学学院客座教授。自 2020 年 7 月起，南开大学统计与数据科学学院教授。2018 年 4 月至今，南开大学“智英健康数据研究中心”主任。科研兴趣集中在数学流行病学以及泛函分析及应用。至今已发表论文 40 多篇，完成一个疫情预警预测平台 EpiSIX 2.0；在 Amer. Math. Soc. 出版专著 Gradient Inequalities: with applications to asymptotic behavior and stability of gradient-like systems，国际期刊 Math. Review 和 Zentralblatt Math 的评论员。新冠疫情以来，向国务院、卫健委、国家疾控中心提交了数百份疫情研判报告，多封报告获得国家领导人批示。

联系方式：huangsenzhong@nankai.edu.cn

代表性成果：

1. **黄森忠**，彭志行，靳祯. 新冠肺炎疫情控制策略研究：效率评估及建议. *中国科学：数学*, 2020, 50(6): 885-898.
2. Bai Y, Peng Z, Wei F, ..., **Huang S**, Xu J. Study on the COVID-19 epidemic in mainland China between November 2022 and January 2023, with prediction of its tendency. *J Biosaf Biosecur* 2023; 5(1): 39-44.
3. Sun Y, Wang M, Wei F, **Huang S**, Xu J. COVID's future: Viral multi-lineage evolution and the dynamics of small epidemic waves without seasonality in COVID-19. *J Biosaf Biosecur* 2023(5): 96-99.
4. **黄森忠**，魏凤英. 新冠疫情趋势预测，在《*新冠再认识*》（徐建国院士主编）第 8 章，共 9 页。
5. **S.-Z. Huang**, A new SEIR model with applications to the theory of eradication and control of diseases, and to the calculation of R_0 . *Mathematical Biosciences*, 2008, 215: 84-104.

计算机视觉与影像分析团队

潘斌：副教授，博士生导师，入选南开大学百名青年学科带头人计划，主要研究方向为图像处理、人工智能。近 5 年，在国际知名图像处理和人工智能期刊 IEEE TGRS、IEEE TIP、IEEE TMM、PR 等先后发表第一作者和通讯作者 SCI 论文 30 余篇，其中 18 篇为中科院一区；先后 4 篇一作或通讯论文入选 ESI 热点论文，6 篇入选 ESI 高被引论文；二十余位国际遥感知名期刊主编对本人的论文进行了正面评价；担任国际知名二区 SCI 期刊《Infrared Physics & Technology》副主编。近 5 年，主持科研项目 13 项，其中国家级项目 5 项，省部级 2 项；研究成果在天津城南变电站、贵州高速、烟台交管等部门获得了应用，取得应用证明 2 项，授权专利 6 项。

联系方式：panbin@nankai.edu.cn

与生物医学合作研究案例：我们与空军总医院开展合作，面向皮肤镜图像，基于深度学习方法，研究了常见皮损的多分类方法。本课题将临床图像和皮肤镜图像相融合，实现不同模态数据间的信息互补与整合，提升了皮损分类的准确率。经空军总医院临床医生鉴定，本课题对典型皮损的分类精度达到 97%。

代表性成果：

1. **Bin Pan**, Zhehao Xu, Tianyang Shi, Tao Li and Zhenwei Shi. An Imbalanced Discriminant Alignment Approach for Domain Adaptive SAR Ship Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2023
2. **Bin Pan**, Qiaoying Qu, Xia Xu and Zhenwei Shi. Structure-Color Preserving Network for Hyperspectral Image Super Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 2022
3. Jiao Liu, **Bin Pan*** and Zhenwei Shi. CR-Famba: A Frequency-Domain Assisted Mamba for Thin Cloud Removal in Optical Remote Sensing Imagery. *IEEE Transactions on Multimedia*, 2025
4. Junzheng Chu, **Bin Pan***, Xia Xu, Tianyang Shi, Zhenwei Shi and Tao Li. Joint Variational Inference Network for Domain Generalization. *Pattern Recognition*, 2024
5. Jiao Liu, **Bin Pan*** and Zhenwei Shi. Cascaded Memory Network for Optical Remote Sensing Imagery Cloud Removal. *IEEE Transactions on Geoscience and Remote Sensing*, 2024

徐君：南开大学统计与数据科学学院副教授，百名青年学科带头人。主要研究方向为计算机视觉和图像处理，目前专注于高效图像修复，编辑和生成。本硕毕业于南开大学基础数学系，2018 年博士毕业于香港理工大学计算机系。2018 年至 2019 年任起源人工智能研究院研究科学家，2019 年 10 月入职南开大学。团队在人工智能、机器学习和计算机视觉领域的国际顶级学术会议 ICML、NeurIPS、CVPR、ICCV、ECCV 和期刊 IEEE TPAMI、IEEE TIP 上发表论文近 40 篇，相关论文的谷歌学者引用一共 5900 余次。团队所研发技术应用于 2020 年武汉新冠肺炎 CT 筛查、360 数科的金融小额放贷业务、魅 KTV 唱功评价系统、PhotoBoost 图像修复 app（全球下载量破百万）。已申请 10 项国家发明专利，5 项获得授权。

联系方式：nankaimathxujun@gmail.com

代表性成果:

1. Coupled Diffusion Models for Metal Artifact Reduction of Clinical Dental CBCT Images. Zhouzhuo Zhang, Juncheng Yan, Yuxuan Shi, Zhiming Cui, Dinggang Shen*, **Jun Xu***. *IEEE Transactions on Medical Imaging*, accept, 2025.
2. Real-World Low-Dose CT Image Denoising by Patch Similarity Purification. Zeya Song, Liqi Xue, **Jun Xu***, Baoping Zhang, Chao Jin, Jian Yang, Changliang Zou. *IEEE Transactions on Image Processing*, vol. 34, pp. 196-208, Dec. 2024.
3. Reducing Fine-Tuning Memory Overhead by Approximate and Memory-Sharing Backpropagation. Yuchen Yang, Yingdong Shi, Cheems Wang, Xiantong Zhen, Yuxuan Shi, **Jun Xu***. *ICML*, 2024.
4. A Dynamic Multi-Scale Voxel Flow Network for Video Prediction. Xiaotao Hu, Zhewei Huang, Ailin Huang, **Jun Xu***, Shuchang Zhou*. *IEEE/CVF Computer Vision and Pattern Recognition (CVPR)*, 2023. (Highlight)
5. Semi-Cycled Generative Adversarial Networks for Real-World Face Super-Resolution. Hao Hou, **Jun Xu***, Yingkun Hou, Xiaotao Hu, Benzhen Wei, Dinggang Shen. *IEEE Transactions on Image Processing*, vol. 32, pp. 1184-1199, Feb. 2023.

自然语言处理与大模型研究团队

李慧嘉：南开大学统计与数据科学学院教授、博士生导师、中国计算机学会高级会员、IEEE 高级会员。主要从数据挖掘、网络分析、群智计算等方面的基础研究工作，并解决生物信息学与系统生物学、大数据分析、人工智能等应用领域的热点问题。主持国家自然科学基金面上、青年项目、北京市自然科学基金面上项目，参与国自科重点及重大研究计划项目多项。已出版中英文专著 4 部，在包括 IEEE 及 ACM 汇刊、PNAS、《中国科学》在内的国内外顶级期刊发表文章 100 余篇，共被引用 3000 余次，10 余篇论文入选“ESI 高被引论文”、“热点论文”及期刊“Featured Article”，成果被图灵奖获得者、中国科学院院士、美国科学院院士等知名专家在 PNAS、Physics Reports 等著名期刊引用评价，并担任《Humanities and Social Science Communications》、《Frontiers In Physics》、《Electronics》、《Frontiers In High Performance Computing》等多个 SCI/SSCI 期刊的编委和学术委员会成员。曾获“北京市优秀青年人才”及首届中央财经大学“龙马学者”青年教授称号，获得英国物理学会高引用作者奖、多个 IEEE 汇刊及 Elsevier 期刊的杰出审稿人等国际学术荣誉。

联系方式： hjli@nankai.edu.cn

代表性成果：

1. **Hui-Jia Li**, Zhen Wang, Jie Cao, Jian Pei and Yong Shi. Optimal estimation of low-rank factors via feature level data fusion of multiplex signal systems. *IEEE Transactions on Knowledge and Data Engineering*, 34(6), 2860-2871, 2022. (CCF A, ESI Highly Cited, “Hot Paper”)
2. Xuelong Li, Marko Jusup, Zhen Wang, **Hui-Jia Li**, Lei Shi, Boris Podobnik, H. Eugene Stanley, Shlomo Havlin, Stefano Boccaletti. Punishment diminishes the benefits of network reciprocity in social dilemma experiments. *PNAS*, 115 (1), 30-35, 2018. (ESI Highly Cited, “Hot Paper”)
3. **Hui-Jia Li**, Haobin Cao, Yuhao Feng, Xiaoyan Li. Optimization of graph clustering inspired by dynamic belief systems, *IEEE Transactions on Knowledge and Data Engineering*, 36(11), 6773-6785, 2024 (CCF A, ESI Highly Cited, “Hot Paper”)
4. **Hui-Jia Li***, Wenze Xu, Chenyang Qiu, et al. Fast Markov clustering algorithm based on belief dynamics. *IEEE Transactions on Cybernetic*, 53(6), 3716-3725, 2023.(CCF B, ESI Highly Cited)
5. **Hui-Jia Li***, Yuhao Feng, Chengyi Xia, Jie Cao. Overlapping graph clustering in attributed networks via generalized cluster potential game. *ACM Transactions on Knowledge Discovery from Data*, 18(1), 27: 1-26, 2023. (CCF B, ESI Highly Cited)

王冀淇：南开大学统计与数据科学学院讲师，研究方向为领域知识增强算法、大语言模型、自然语言处理与时空数据挖掘。其核心研究聚焦于领域知识表示与融合方法，通过结构化领域知识提升通用模型在垂直场景中的性能。在法律 AI 领域，提出的领域知识增强算法使中小模型在裁判文书分类、摘要生成及法律推理任务上达到甚至超越一般大模型性能；在时空数据挖掘方向，围绕下一兴趣点推荐、异常轨迹检测等任务实现了一定突破。相关成果发表于 ACL、NAACL、ECAI 等国际顶会，推动了领域知识在人工智能模型中的高效应用。

联系方式：qiqi.wang@nankai.edu.cn

代表性成果：

1. **Qiqi Wang**, Ruofan Wang, Kaiqi Zhao, Robert Amor, Benjamin Liu, Jiamou Liu, Xianda Zheng, Zijian Huang, "SKGSum: Structured Knowledge-Guiding Document Summarization", In *Findings of the Association for Computational Linguistics: ACL* 2024.
2. **Qiqi Wang**, Ruofan Wang, Kaiqi Zhao, Robert Amor, Benjamin Liu, Xianda Zheng, Zijian Huang, "Towards Legal Judgment Summarization: A Structure-Enhanced Approach", In *26th European Conference on Artificial Intelligence (ECAI)*, 2023.
3. **Qiqi Wang**, Kaiqi Zhao, Robert Amor, Benjamin Liu, Ruofan Wang, "D2GCLF: Document-to-Graph Classifier for Legal Document Classification", In *Findings of the Association for Computational Linguistics: NAACL* 2022.
4. **Qiqi Wang**, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao, "DRVC: A Framework of Any-to-Any Voice Conversion with Self-Supervised Learning", In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

张瑶：讲师，主要致力于自然语言处理、知识推理和对话系统的研究，获得天津市科技进步二等奖一项，主持国家自然科学基金项目。以第一作者或者通讯作者在 ACL、EMNLP、AAAI 和 SIGIR 等国际顶会及期刊发表论文十余篇。在知识推理方向，创新提出三元关系建模方法与动态多跳推理框架，突破传统二元关系局限，显著提升长距离推理效率；在多模态研究中构建跨模态语义对齐技术，提出视觉区域集合操作理论，形成端到端多选问答生成范式；在对话系统领域开发知识预选机制与医疗可信度评估模型，有效提升对话相关性与安全性。

联系方式：yaozhang@nankai.edu.cn

代表性成果:

1. Qin L#, **Zhang Y**#, Liang H, Wang J, Yang Z*. Well Begun Is Half Done: Generator-agnostic Knowledge Pre-selection for Knowledge-grounded Dialogue. *EMNLP*, 2023.
2. **Zhang Y**, Li P, Liang H, Jatowt A, Yang Z*. Fact-Tree Reasoning for N-ary Question Answering over Knowledge Graphs. *ACL (Findings)*, 2022.
3. Lei W#, **Zhang Y**#, Song F, Liang H, Mao J, Lv J, Yang Z*, Chua T S*. Interacting with Non-Cooperative User: A New Paradigm for Proactive Dialogue Policy. *SIGIR*, 2022.
4. **Zhang Y**, Zhang X, Wang J, Liang H, Lei W, Sun Z, Jatowt A, Yang Z*. Generalized Relation Learning with Semantic Correlation Awareness for Link Prediction. *AAAI*, 2021, pp.4679-4687.
5. **Zhang Y**, Liang H, Jatowt A, Lei W, Wei X, Jiang N, Yang Z*. GMH: A General Multi-hop Reasoning Model for KG Completion. *EMNLP*, 2021, pp.3437-3446.