# Science Ascend
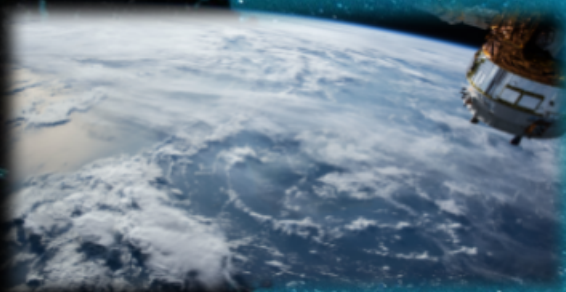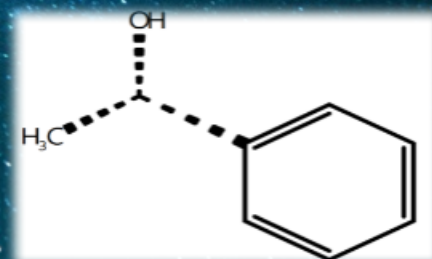
*Rising to new heights of discovery with Science!*
*Every week!*

Origin of chiral asymmetry, ALMA, ZTF, ultracool dwarfs, JWST, asteroids and more!

Charge-detection MS, high-entropy alloys, PFAS detection,reaction kinetics with spectroscopy, microfluidic sensors and more!

Sentinel-1 and 2 fuse, agent mining, thermal + lidar + RGB fuse, better UAV monitoring and more!

Indoor PM measurement insights, SVHC replacement works, mobility of selenium and more!

cout<<solutions;

Much better theoretical backgrounds, robust forecasting, efficient explainable AI, better optimizers and more!

# Science Ascend

*Rising to New Heights of Discovery!*

*Science Ascend* teleports you to the frontiers of science. It compiles and discuss the scientific research preprints from arXiv, bioRxiv, chemRxiv just from the previous week to be cognizant of the *state-of-the-art* of knowledge in astrophysics, chemistry, environmental chemistry, remote sensing, and applied statistics/data science. Light from the *Science Ascend* will keep brightening the dark horizon beyond the limits of our comprehension. FIRE Araştırma Eğitim Ltd. Şti. guarantees the weekly publication and dissemination of this journal, and make it available for everyone at most fifteen days after its publication freely.

# Bilim Yükselişi

*Keşfin Yeni Yükseklerine Ulaşmak!*

*Science Ascend* sizi bilimin sınırlarına ışınlar. Astrofizik, kimya, çevre kimyası, uzaktan algılama ve uygulamalı istatistik/veri bilimi alanlarındaki bilgi birikiminin *en son durumu* hakkında bilgi sahibi olmak için arXiv, bioRxiv, chemRxiv'den sadece bir önceki haftaya ait bilimsel araştırma ön baskılarını derler ve tartışır. *Bilim Yükselişi*'nden gelen ışık, kavrayışımızın sınırlarının ötesindeki karanlık ufku aydınlatmaya devam edecektir. FIRE Araştırma Eğitim Ltd. Şti. bu derginin haftalık olarak yayınlanmasını, dağıtılmasını ve yayınlandıktan en geç on beş gün sonra ücretsiz olarak herkesin erişimine açılmasını garanti eder.

# Last week in Astrophysics

*Author: Yasin Güray Hatipoğlu*

The preprints summarized here were published between September 24 - September 30, 2024. These are from arXiv's astro.EP cross-fields without high-energy main cross-list papers.

## Interstellar Medium/Galaxy

Hoang[1] stated that chiral asymmetry might be induced by spin-polarized electrons (SPEs), whose spin-polarization originates from the Barnett Effect[1] and are originated from magnetically-aligned dust grains after the exposure to interstellar ultraviolet (UV) irradiation. Rotation comes from the gas collisions and radiative torque (RAT)[2] Dust grains are silicates with iron embeds, which impart superparamagnetism. The other significant part of the paper illustrates how these SPEs can induce mono-chirality by showing different potential energies of interactions while considering the direction and spin of the electrons with the molecules.

## Stellar Systems - Populations - Clusters

Hsieh et al.[2] utilized the Atacama Large Millimetre/Submillimetre Array (ALMA) for studying protoplanetary disks and monitored the following nearby clouds: Corona Australis, Aquila, Chamaeleon I & II, Ophiuchus North, Ophiuchus, Serpens. The entire sample was self-calibrated with auto_selfcal, with several modifications including a manual mask input instead of automasking from tsclean. The source identification was made with SciPy's ndimage.maximum_filter[3] with the same filter size as the beam size. They used three weighting schemes, natural (all pixels have the same weight, poor side-lobe levels, but better noise performance), uniform (inversely proportional to the antenna-related sampling density function, higher noise than natural), and Briggs 0.5 (a compromise between uniform and natural weightings)[4], in protostellar dust disk radii measurements. They detected 184 protostellar disks

---

[1] An uncharged body starts to rotate and along its axis, the electron spins aligned with the induced magnetic moment in the opposite direction.

[2] The radiation hits the irregularly shaped body anisotropically and induces some rotational velocity.

[3] An image operator filter where one sets a filter window and every pixel within this window will be assigned the maximum value present in the filter window.

[4] Further information on weighting schemes can be found here.

(18 of them are new, and 7 of them had the first millimetre measurements) and reported size differences among protoplanetary disks while moving through the stellar evolution ladder, effectively shrinking with older ages.

## Single Star System (Star, Exoplanet)

Voloshina et al.[3] utilized Zwicky Transient Facility (ZTF) data release 8 for M-dwarf flare detection. For this purpose, they applied two different methods: 1) **parametric fit search**, and 2) **machine learning**. For the **parametric fit search**, continuous data chunks with the highest cadence were separated with a minimum duration of 30 minutes and 2 hours separately as long- and short-duration chunks. They also applied $\chi^2$ statistics-filter as value greater than 11 to only work on variable ones. In the end, they had around 4 million long-duration and 10 million long-duration chunks. Then, they defined a modified analytical function to model stellar flux density and fitting was done in light-curve and iminuit packages in the Python environment. For flares with sufficient data points, light-curve OtsuSplit feature was used, and in the end, they had 308 candidate flares. For the **machine learning**, they employed Active Anomaly Discovery algorithm, which is tuned by experts, and the experts chose to limit the anomalies for M-dwarf flares, and out of 860 objects examined, 35 objects were short-listed for further analysis. Flare energies were calculated for the sample with lower than $< 20\%$ parallax uncertainty from Gaia EDR3. After selecting the ones meeting this criterion and having enough data points, they ended up with 13 flares. They assumed a 9000 K optically thick black body for flares and always used the r-band of the ZTF. They determined the spectral class photometrically from Pan-STARRS DR2 after galactic reddening correction, and then, for the extinction values. In the end, both methods have inputs preprocessed differently, and also resulted in outputs that can complement each other.

Brooks et al.[4] with the Backyard Worlds: Planet 9 Collaboration utilized machine learning techniques and discovered 118 new ultracool dwarf candidates. They mainly utilized Wide-field Infrared Survey Explorer (WISE) data, and to be able to detect even fainter objects in the second band of the WISE, W2, they used the SMDET convolutional neural network technique. WISE images were co-added to each other in un-WISE project and the resulting time series were worked on in the training step of SMDET with synthetic objects. They were looking for fast-moving objects to detect relatively nearby brown dwarfs, hence in the time-series data, the esti-

mated location-pixel for the object of interest was also critical, which the algorithm tried to predict. In another part of the same study, via Backyard Worlds: Planet 9 citizen science project, humans checked the images for the candidate high proper motion objects and marked 1730 real candidates out of 11900 SMDET candidates. Among these 1730 targets, 232 of them were not cataloged according to the SIMBAD and VizieR Catalog Access Tools. Finally, Gaia DR3 detected objects were also filtered out since the target ultracool dwarf samples were not expected to be visible in Gaia bands (visible and near-infrared). The photometry data they compiled were from WISE W1 W2, AllWISE W3 W4, J and $K_s$ of VISTA Hemisphere Survey, UKIRT Hemisphere Survey J and $K_{MKO}$ bands, and PanSTARRS DR2 bands of $g_{ps}$, $r_{ps}$, $i_{ps}$, $z_{ps}$, and $y_{ps}$. They validated the spectral classification from the photometry data on two objects with sufficiently accurate estimations. They also stated the benefits of utilizing SMDET in working with large-scale datasets for similar discoveries.

Li et al.[5] specifically worked on the M3.0V spectral type star G 80-21 from the Calar Alto high-Resolution search for M dwarfs with Exoearths with Near-infrared and optical Échelle Spectrograph (CARMENES). They chose this star as it has high chromospheric activity. To model the quiet regions of G 80-21, a similar but inactive star from the CARMENES catalog was chosen (LP 819-17). They constructed the model atmosphere of G 80-21 with MARCS code (line blanketed local thermodynamic equilibrium (LTE) model grids). On top of the photosphere, they put lower and upper chromosphere, and transition region, and after calculating them under hydrodynamic equilibrium. The MULTI program computed the non-local thermodynamic equilibrium for the whole atmosphere and along with the magnetic fields, they were the initial conditions for the RH1.5D code. After these, the active model was iteratively being matched with the actual observational spectra of the G 80-21 by the visual examination of the researchers. Two active regions in conjunction with an inactive region was the optimal fit and the active region was variable and could even take up to 82 % of the sphere.

Tu, Wang, and Liu[6] utilized James Webb Space Telescope (JWST) Near Infrared Spectroscopy (NIRSpec) and Mid-Infrared Instrument Low Resolution Spectroscopy (MIRI LRS) to analyze 20 T and Y dwarfs. They used ATMO2020++ and Sonora Elf Owl atmospheric model grids, and they also compared the alternatives of using NIRSpec + MIRI LRS together or separately. The data normally were reduced by the standard jwst pipeline, while for photom-etry calculations, both the JWST pipeline and CRDS, and also DOLPHOT 2.0 software were used. Among other things, they discussed obtaining effective temperatures from two different spectra in this study. They stressed the importance of which instrument is sensitive to which molecules and lines according to their working region. Nevertheless, MIRI LRS only performed poorly compared to the combined or NIRSpec-only ones.

Vázquez et al.[7] used Gaia DR3 and artificial intelligence techniques to classify hot subdwarf binaries in large datasets. The Support Vector Machine was their supervised classification for the 3084 hot subdwarfs according to their color-magnitude diagrams, and self-organizing maps (SOM) and convolutional neural networks (CNN) were also used for the 2815 of them with Gaia DR3 BP/RP spectra. Before this SOM - CNN analysis, a pre-analysis of the spectra (each has 308 bins) took place with either the uniform manifold approximation and projection (UMAP) or the cosine similarity. Their results had a high-agreement level with the Virtual Observatory SED Analyzer (VOSA) tool's binaries.

### Protoplanetary/Circumstellar Disks

Hu et al.[8] worked on Bernhard-2, a binary system with a misaligned circumbinary disk. They obtain spectra from Gran Telescopio CANARIAS OSIRIS (Optical System for Imaging and low-Intermediate-Resolution Integrated Spectroscopy) instrument, and also CoRot-7 for testing the radial velocity (RV) measurement stability, all in the long-slit mode with R2500R grism and the slit width of 0.6". The data were reduced with the PypeIt package, the sky spectrum and telluric lines (lines coming from emissions from Earth-residing species) were removed by airvacuumvald package. The wavelength-calibrated spectra was cross-correlated with a single-star spectrum template using iSpec and Phoenix packages, too. Two more spectra for Bernhard-2 were obtained via Keck Planet Finder on the Keck I telescope and the spectra were reduced with its specific data reduction pipeline. Additionally, a Magellan Folded-port InfraRed Echelette (FIRE) spectrum was also obtained for Bernhard-2, which was in the infrared region of the light spectrum. Lastly, the photometric observations were obtained in the expected occultation time with Post Observatory in r and i bands. In the end, they confirmed that the circumbinary disk is indeed misaligned with around 0.69 eccentricity and the Bernhard-2 binary is a confirmed KH 15D-like system.

Williams et al.[9] published the Planetary En-

riched White Dwarf Database[5] In September 24, 2024 the database contained 1739 white dwarfs, and the preprint has discussed the metal enrichment, and in case there is a binary companion within 200 au the enrichment is suppressed.

## Sun - Solar System

Eya et al.[10] focused on automated data analysis of Forbush decreases[6]. They utilized the daily-averaged cosmic ray raw data corrector for atmospheric pressure and also solar activity-related data (solar wind speed, Dst index etc.) from the Omni web interface of NASA. The data came from the neutron monitor stations and the measurements were between 1998-2006. A substantial amount of work was on relating the event records between different stations since they were expected to be *the same* due to the global nature of Forbush decreases. They confirmed this and also related the Forbush decreases to solar wind speed and other parameters. As a future work, they declared that they will work on separating the cosmic ray diurnal anisotropy-related variations in the data.

Li et al.[11] constructed a multi-station meteor monitoring and tested two stations in the suburbs of Beijing with 0.3-0.4 arcmin accuracy during the Geminid meteor shower. The multi-station system, with wide field-of-view lenses and CMOS cameras, detects the meteors when they enter the Earth atmosphere with one microsecond accuracy. They will further develop this construction and increase the station numbers as well.

---

[5]The related Python Package GitHub repository is here.
[6]Forbush decrease is the reduction in observed galactic cosmic ray intensity since a solar wind after a coronal mass ejection (CME) event scavenges some of them.

# Last week in Chemistry

*Author: Yasin Güray Hatipoğlu*

The preprints summarized here were published between September 24 - September 30, 2024. They are more in nature of spectroscopy alone, and hence several studies regarding biochemistry, chromatography, and several other disciplines might be missed here.

## Mass Spectroscopy

Sanders et al.[12] worked on a multi-injection online size exclusion chromatography and charge-detection mass spectroscopy (CD-MS) system which also included Hadamard transform multiplexing. The multi-injection online is a self-explanatory name, a bit-index injection plan was set and multiple injections were done from the sample, and the Hadamard transform multiplexing considers all these injections simultaneously and demultiplex these signals in the end. The size exclusion chromatography separates the compounds according to their sizes. Mass Spectroscopy is, basically, ionizing a molecule/atom and after monitoring its movement through the magnetic field and determining the mass/charge ratio.[7]. The charge-detection version of mass spectroscopy also gives an output regarding the charge of the molecule or molecular part where the information of m/z is known. As a result, the mass can be estimated **directly**. They worked on $\beta$-galactosidase, GroEL, and *E. coli* cell lysate, and were satisfied with the results.

Laurent et al.[13] illustrated the utility of time-of-flight secondary ion mass spectrometry on monitoring high-entropy alloys. They worked on Ru-Pt-Pd-Ir-Rh alloy. The idea to use ToF-SIMS here is that the ion cluster it generates might give the information on the neighboring atoms around a specific atom, hence facilitate the homogeneity assessment. They utilized principal component analysis in finding the most effective ion to generate secondary ions from the Ru-Pt-Pd-Ir alloy and $Bi_5^+$ was chosen. One important part of this study was the automatic cluster peak identification and processing. Several of the elements they worked on had multiple isotopes and comparatively complicated m/z spectrum. They created a matrix that can accomodate all species worked in the study with their corresponding isotops as different non-zero cells in the matrix. This is then used to create an expected simulated spectrum from the high-entropy alloy sample. The real sample also has an error matrix in their measurements and to minimize this error in retrieving the mixing ratio in the alloy, a method based on the non-negativity-constrained linear least squares regression[14] was employed. They reported that their approach was highly beneficial in this domain.

Hua et al.[15] employed metal-organic frameworks (MOF) to enhance per- and polyfluoroalkyl substances (PFAS). They first solid-phase extracted the PFAS compounds with the frameworks and then measured them with the liquid chromatography tandem mass spectromety (LC-MS/MS). Among the six different MOFs they utilized, UiO-66[8] was the best with the 87% adsorption and 85% recovery efficiency for 33 PFAS molecules.

Weber et al.[16] reported the high-resolution upgrade of Nano Secondary ion Mass spectrometry - NanoSIMS. They reported the updates on $Cs^+$ source, high voltage control, stage reproducibility, and other parts. They presented the results of the analyses on aluminium samples containing silicon crystals, microalgae, and fungi-colonized plant roots, with down to 30 nm lateral spatial resolution.

## Fluorescence Spectroscopy

Nayak and Wilson investigated the cation effects on uranyl-nitrate[9] complexation equilibria with the time-resolved laser-induced fluorescence spectroscopy. They reported that sodium, lithium, and ammonium cations mostly speciated the U(VI) into $UO_2^{2+}$ and $UO_2(NO_3)^+$, but tetramethylammonium cation resulted in different, potentially higher-order complexes.

## Infrared Spectroscopy - IR

Freitag et al.[17] worked on the wheat quality assessment by mainly utilizing near-infrared spectroscopy data. They combined analysis of variance simultaneous component analysis[10] and near-infrared spectra to examine annual and regional impacts on wheat. They used a benchtop spectrometer in the 400 - 2499.5 nm range to

---

[7]From then on, the interpreter should work a way through finding out the actual mass. This is no different from a complex puzzle, and the complexity increases with the size of the molecules. There are many reference spectra and databases to aid in this, some are public, some are paywalled.

[8]Universitetet i Oslo, $[Zr_6O_4(OH)_4]$ clusters, more information with a 3D model can be found here.

[9]This specific complexation is of interest as the spent nuclear fuel is treated with nitric acid.

[10]ASCA is like a combination of analysis of variance and principal component analysis, where we are only checking for the effect of two different things and their combined effect. Further information can be found here[18]

measure the near-infrared spectrum, too. They reported that the sampling site and year of collection significantly vary the spectra, and also the protein, starch, moisture, fat, fiber, and ash content of the wheat.

## Nuclear Magnetic Resonance - NMR

Soussi-Therond et al.[19] utilized the Nuclear Magnetic Spectroscopy and several concepts regarding that to monitor glucose-6-phosphate (G6P) dehydrogenase's reaction mechanism. Especially with the dissolution dynamic nuclear polarization (dDNP), hyperpolarizing magnetically active nuclei in cryo temperatures, the researchers reported the G6P oxidation by the G6P dehydrogenase and delineated the boundaries of this technique in studying kinetic models and hypotheses. NMR peaks in different seconds after the start of the reaction provides critical details of the reaction pathway.

## Microfluidic Sensors

Zhang et al.[20] developed a microfluidic sensor to greatly speed up the motile plant zoospore detection with a portable device. They made COMSOL simulations on how the electrical field would response, and also utilized chemotaxis to lure the spores to the detector. Their devices were constructed with photolithography. They were successful in detecting zoospores with this method.

# Last week in Remote Sensing

*Author: Yasin Güray Hatipoğlu*

The preprints summarized here were published between September 24 - September 30, 2024. These are generally based on the preprints retrieved when "remote sensing" words are given between quotation marks within arXiv's cs.CV and similar cross-fields.

## Segmentation

Russo et al.[21] fused Sentinel-1 and Sentinel-2 data for powerful reservoir monitoring. Their end-to-end deep learning framework included a spatiotemporal datacube of synthetic aperture radar (SAR) polarization, elevation, slope, and Sentinel-2 multispectral bands. The framework (SEN12-WATER) removes speckle noise from SAR, segments water bodies with a U-Net architecture, conducts time series analysis, and includes Time-Distributed-Convolutional Neural Network (TD-CNN) within the Keras-TensorFlow Python package. Their algorithm was verified through the result comparison with the ground truth. Their training included both binary cross-entropy loss and GapLoss function. They also aimed to predict the next two months' water mask in a frame using the previous 14-month data with ConvLSTM, Bidirectional ConvLSTM, and Time-Distributed CNN. Their segmentation was very successful, and slope, elevation, Sentinel 2 10-meter spatial resolution bands, and both Vertical-Vertical and Vertical-Horizontal polarizations of S1 together produced the highest accuracy. As for the next frame prediction, TD-CNN had the highest scores in all three metrics.

Bi et al.[22] constructed the Agent Mining Transformer - AgMTR, a few-shot segmentation transformer to alleviate problems arising from intra-class variability and problematic background in remote sensing. The method contains three parts: The first part is the Agent Learning Encoder, where the agents learn the important parts of a specific object, e.g., a plane, separately (wing, tail, etc.). Then, the Agent Aggregation Decoder explores the semantics of the unlabeled data and guides the agents to classify-get enhanced over their targets with this information. The last part, Semantic Alignment Decoder works on the target image, and basically, it assumes that the same objects are more similar to each other in terms of pixel-level similarity, and it constructs the pseudo-local mask of the target/query image. They validated this novel approach on the iSAID dataset, and also on the PASCAL-5 and COCO-20 datasets. They outperformed the R2Net (ResNet-50 and ResNet101 backbones) and FPTrans on the iSAID, PASCAL-5, and ResNet-50, ResNet-101, FPTrans, and MuHS on the COCO-20 datasets, too.

Gordon et al.[23] worked on rhino communal defecation sites, middens identification from 9722 remote sensing imagery (heat-thermal, RGB, and lidar) of Kruger National Park. They approached this issue through developing an active learning method[11]. They transferred the learned model from the VGG16 model trained on the ImageNet dataset. The model used either thermal, RGB, or lidar, or several combinations of them (fused). Firstly, they passive learned the data and tested 20% of it, and thermal + RGB and Thermal + Lidar fused sets were the best in accuracy, precision, recall, and F1 scores. Then, they chose to report thermal + RGB fuse in active learning as well, with very similar test accuracies but drastically less data requirement (9736 vs 500 images in training) and runtime (81 hours vs. approximately 4 hours).

## UAV

Zhang et al.[24] described PolRa - Portable L-band radiometer to mitigate the coarse spatial resolution problem in remotely-sensed soil moisture products, such as Soil Moisture Active Passive (SMAP) outputs. PolRa, instead, can be mounted on towers or UAVs and remotely retrieve the soil moisture information in sub-meter resolution. While having the ground truth with the handheld Time domain Reflectometry (TDR) sensor, they also retrieved PolRa brightness temperature data and estimated the soil moisture from it with several algorithms, and compared the match with the ground truth by TDR. The authors stated that in very fine spatial resolution, PolRa reliably captured the temporal variations and the unbiased root mean square error values were lower than 0.04 $m^3/m^3$.

## Modelling-Forecast

Guzman-Lopez et al.[25] stated that they merged the agricultural census data and remotely-sensed data products with sparse regression and Elastic-Net regularization, and established Granger predictive causal[12] relationships between paddy rice

---

[11] Several labeled samples are given to the model. Then for the following cases where the model would be the most uncertain, the labeler is queried by the model to label them, so that with the least amount of labeled images the concept can be learned.

[12] Being able to predict a value of a parameter using another parameter's earlier value.

yields and aforementioned data proucts. The remote sensing data came from Moderate Resolution Imaging Spectroradiometer (MODIS) sensor's MOD13Q1 NDVI data, the precipitation from the CHIRPS Pentad, and the temperature was from the MODIS MOD11A1 data product[13]. They also took the first and second derivatives of these predictors and included them as velocity and acceleration components, such as VEL_NDVI. Then, regression with elastic-net, gradient tree boosting, and generalized additive models were employed to predict agricultural yield. The mean square error metrics favored XGBoost, yet, quite low training error, and almost 4-times of it in the validation was hinted at potential overfitting, and elastic-net regularization was chosen to be best with a slightly lower validation error.

## Object Detection

Zhao et al.[26] proposed the OrientedFormer[14], an end-to-end object detector for remote sensing. The structure is a backbone and a decoder (consecutive application of the self-attention, cross-attention, and feedforward-feed network). For the object query, there were separate content queries and positional queries. For the positional encoding, rather than commonly used horizontal boxes, they used Gaussian positional encoding (PE). The datasets they utilized were DIOR-R, DOTA v1.0, 1.5, 2.0, HRSC2016, and ICDAR2015. OrientedFormer mostly outperformed other algorithms on all datasets.

---

[13]Readers should be cautioned that MODIS is **NOT** a satellite but a sensor, and MOD11A1 or MOD13Q1 are **NOT** sensors, but data products.

[14]The code can be found in this GitHub repository.

# Last week in Environmental Chemistry

*Author: Yasin Güray Hatipoğlu*

The preprints summarized here were published between September 24 - September 30, 2024. chemRxiv's Earth, Space, and Environmental chemistry preprints are being surveyed, and unfortunately, not many preprints are published under environmental topics in this field.

Adelodun et al.[27] measured the impact of the sensor locations on indoor air quality monitoring in a controlled, 82.2 $m^3$ room. There were three different sensors, one directly at the inlet of the air purifier, another one on the wall, and another one at a typical breathing height at the center. Particulate matter sources were incensing, and four of them were placed in different locations and they were opened one by one. Then, Particulate Matter 1 (PM1, particles smaller than 1 $\mu m$), and PM2.5 were measured. The wall sensor measurements were always lowest in the PM measurements, potentially, *inter alia*, because of its further distance to the polluters, and the other two sensors were also not completely the same as each other.

London, Glüge, and Scheringer[28] reported a multi-criteria decision analysis to replace banned-restricted chemicals, with specific focus on the European Union regulation on the Registration, Evaluation, Authrosiation and Restriction of Chemicals (REACH) and its substances of very high concern (SVHCs) list[15]. Their method was based on Multi-Attribute Value Theory with discrete value functions. On top of the SVHCs criteria, they also add mobility, high ecotoxicity, global warming potential, and ozone depletion potential. Their dataset had two different parts: one hypothetical and one real. The hypothetical one has 256 combinations of persistence, bioaccumulative, human and ecological toxicity with very high, high, moderate, and low options (148 out of 256 have the SVHC characteristics). The real one was from the 16 alternative chemicals to the decabromodiphenyl ether, a persistent organic pollutant brominated flame retardant. The coefficients were arranged to ensure a good value does not act like a trade-off for a bad value, and as a result, having a very highly hazardous trait penalizes the candidate more than what would have been without. Compared to the GreenScreen benchmark as well, their decision criteria method placed more emphasis on low ecotoxicity rather than persistence or bioaccumulation.

Laberge-Carignan et al.[29] studied the impact of temperature on selenium mobility under different reductive-oxidative conditions. They constructed flow-through reactor experiments to check the impact of various parameters on selenium mobility under environmentally relevant conditions and selenium concentrations. There were aged ($> 50$ years old) and fresh ($< 6$ years old) organic mater, 4 or 21 degree Celsius temperature, 7, 70, or 100 nM [Se], and $HSeO_3^-$ or $SeO_4^{2-}$ selenium-containing species. Then, they measured the outflow with inductively-coupled plasma atomic emission spectrometry[16]. They measured the outflow for several minutes and created temporal profiles. They reported that fresh organic matter species were more effective in removing selenium, and different selenium species have contrasting outflow rates under different temperatures.

---

[15]This is related to the Article 57 of the REACH, which contains a) carcinogenicity, b) mutagenicity, c) toxicity against reproduction, d) persistent, bioaccumulative, and toxic, e) very persistent and very bioaccumulative vPvB, f) endocrine disruptiveness, g) probable serious harmful effect to human health and nature other than listed above but still nonnegligible.

[16]For example, the argon gas heated up to 12000 degrees celsius to transit into a plasma state, the sample is nebulated to this plasma and the emissions from the atoms originated from the plasma treatment of the sample are measured in the end.

# Last week in Data Science- Applied Statistics

*Author: Yasin Güray Hatipoğlu*

The preprints summarized here were published between September 24 - September 30, 2024. This is generally from arXiv's stat.ML (and correspondingly cs.LG) cross-list. Large-language model-related, or text-mining and similar studies are omitted, also several less-application oriented studies (they, too, are important, but currently *Science Ascend* can't accomodate to review them within the time constraints.

## Denoising

Fan et al.[30] made a curious study on clustering the speckle noise with their speckle unsupervised recognition and evaluation (SURE) algorithm. They used a 523 nm continuous wave originated from a laser generator. They experimented this technique in glucose sensing and harsh-condition multimode fiber (MMF)-based communication. For MMF dynamic envronment, 2 meter multimode fiber was wound into coils with a 10 cm diameter and an example environmental disturbance was simulated by a motorized rotator with a 20 degrees/s constant velocity and two loose clamps for random friction application. For glucose, glucose stock solutions were diluted to two different concentration relevant for a healthy and hyperglycemic people, respectively. For the speckle image clustering task, MNIST dataset with bicubic interpolation from the 28x28 size to 400x400 was utilized. On the clustering part, SCAN and SHACK were implemented, where the former has a convolutional neural network for feature extraction and a fully connected layer for clustering and an overclustering, and the latter is an hierarchical agglometarive clustering. The results were satisfactory.

## Time Series

Katende[31] constructed a framework to impute missing data and analyze the structural transformation (the shift from a dominant agrarian economy to a diversified industrial-service economy) for country-sector-year dimensions economy-related values with low-rank matrix assumption, Bayesian hierarchical modelling, and machine learning - Least Absolute Shrinkage and Selection Operator (LASSO). Bayesian Hierarchical Modelling had the following main equation for each country-sector element =

$$y_{it} \ N(\mu_{it}, \sigma^2)$$

and for the $\mu_{it}$:

$$\mu_{it} = \beta_0 + \beta_1 X_{it} + \gamma_i + \delta_t$$

where $\beta_0$ is the intercept and $\beta_1$ is the slope of $X_{it}$, where this is the covariates, such as public investment, where these can be relevant for each country-sector, $\gamma_i$ is for the random effects, and $\delta_i$ is global trends. They reported better results of this approach in their simulation study, and its high predictive accuracy persisted even with only 40 % of the data available.

Elborough et al.[32] devised a way to compute Shapley values[17] for time-series-like data, just as in image processing super-pixel cases to not overload the resources. Their sample case was forensic DNA classification. Their Shapley values were region-block-like rather than for every value-allele-gene. The base model was a convolutional neural network, constructed in R v.4.2.3 with the Tensorflow, Keras, dplyr, and SOAR libraries. The Shapley value was calculated as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(x_{S \cup \{i\}}) - f(x_S)]. \tag{1}$$

The way to calculate these values to the blocks was the Kernel SHAP algorithm. They iteratively shrank the part to calculate Shapley values by starting from each dye lane, selecting the largest Shapley value, and partitioning these parts, and repeating the process. In this way, less Shapley computation is required, but the result might still be quite specific. They were successful in making this case rapid and explainable.

Caljon et al.[33] studied the impact of dynamic loss weighting applied in the training phase of time-series forecast models on both accuracy and stability. In other words, they looked for ways to increase forecast stability without undermining the accuracy. Their approach, Task-Aware Random Weighting was found to be the best among other methods. Forecast stability means the difference in the forecasted value of the same time-step for two consecutive estimations. The total loss estimation considers both the error in the observation and the stability. For instance, in the case of the N-BEATS-S network, there is a static coefficient on putting a weight for these losses, and the error term considers both the error in previous occasion and the present occasion, while the stability considers the difference in the forecasted value of previous occasion and present occasion. Finally, dynamic loss weighting changes this static coefficient to a dynamic one, essentially putting

---

[17]A way to increase the explainaibility of artificial intelligence.

different weights on the model accuracy and stability while calculating the loss. Among the different options to tune this, their way was randomly sampling a coefficient from a uniform distribution that is always smaller than $\kappa$, which is the tunable hyperparameter and can prevent excessive prioritization of the model stability over its accuracy. Their algorithm was better than GradNorm, Uncertainty Weighting, Random Weighting, Gradient Cosine Similarity, and Weighted Gradient Cosine Similarity.

Marx, Kuleshov, and Ermon[34] worked on outlining a robust forecast uncertainty reporting framework even in the presence of unpredictable distribution shifts, feedback loops, and adversarial actors. They made use of the Blackwell Approachability and game analogies and generated practical algorithms as well, and reported that in the end, the forecast and calibration for the energy systems can be improved with their strategy. Qin et al.[35], too, considered the issue of distribution shifts in the time series data and proposed the Evolving Multi-Scale Normalization framework[18]. The first reason to do this was the multi-scale different variations in reality, such as annual or weekly variations. Periodicity extraction utilizes the Fast Fourier Transformation (FFT) technique. The following step's slices-windows to create statistics will be prepared according to the periodicities extracted by the previous step. In other words, a "normalization" will take place, then, a denormalization will try to model the non-stationarity with a backbone forecasting model after concatenating the denormalized slices. In the last multi-scale adaptive ensemble, the local periodicity is also calculated by the FFT. The EvoMSN improved many different backbone forecasters in almost all cases and metrics, and it was the best among the online learning strategies Online-TCN, FSNet, Experience Replay, DER++. The EvoMSN was also the best in most cases considering the normalization methods SAN, RevIN, Dish-TS.

## Machine Learning / Deep Learning

Hoang[36] considered the poison of dimensionality, very-high model size-originating problems in the machine learning, as well as posion coming from the data, where especially in social media and human-related data, the assumption of "data is honest" is unrealistic. The author aimed to check the impact of the model size on the security of the model with changing the number of honest (H) and poisoned (P) data. They stated that if the number of dimensions is higher than $169H^2/P^2$, the security is impossible, and they

worked on the ones smaller than that size and presented the poison data impact on the model. They also considered dimension reduction and also stated that the number of dimensions for the optimum case is related to the H/P.

Wang and Li[37] focused on better theoretical and practical comprehension for the out-of-distribution deviation-shift mitigation in machine learning applications. They stressed the importance of out-of-distribution generalization, such as being trained in birds in cages, then, being able to classify birds in forest as birds, too. Moreover, novel semantics may also be encountered, like no training on dogs, and getting images including dogs, and these should not be misclassified with something from the model's training data. They approached this issue with a graph-theoretic framework and utilized singular value decomposition to understand the impact of these wild data on the model. They found this is better than the state-of-the-art *Scone*.

## Modelling

Paul et al.[38] worked on parameter-efficient fine-tuning (PEFT)[19] methods for Depth Anything vision foundation model's post-hoc Bayesian inference. Their domain was monocular depth estimation. They tried the Stochastic Weight Averaging Gaussians (SWAG), BitFit, DiffFit, and CoLoRA. They conducted experiments with the DINOv2 and DPT encoders, worked on continuing fine-tuning with the checkpoint records, and used the NYU and KITTI datasets. Especially on the KITTI dataset, the CoLoRA approach was the best choice.

Bourdais and Ohwadi[39] studied the model aggregation and how it can be optimized via minimal error aggregation and minimal variance aggregation metrics. They took the models as black boxes of input-output functions without assuming much about their nature. They chose to aggregate models in a point-wise linear manner, and they also presented examples of how simple-averaging on the individual utility of the models is far from the real case. In the end, they showed the outperformance of minimizing variance compared to minimizing empirical error.

Wang, Yin, and Li[40] considered a debiased semi-supervised learning training phase while finding ways to reduce the confirmation bias with the TaMatch framework. They reported that batch sizes should be large enough to reduce the bias amplification. A critical detail is strong and

---

[18]The GitHub repository for the EvoMSN is here.

[19]This is an approach to still be able to tune very-large, very high number of parameter-including models even while using modest computing resources. It only tunes some of the parameters, hence preventing catastrophic forgetting and overfitting issues.

weak classes, where the truth probability of a class is lower than the expected marginal distribution of the model or vice versa, the class is defined as strong and weak, respectively. The debiasing is expected to happen if the strong class is suppressed and the weak class is augmented. TaMatch mostly outperformed many other algorithms in CIFAR-100, STL-10, and EuroSat datasets, and it was also better in several other aspects.

Jiang, Lu, and Willett[41] worked on the simulation-based inference method Embed and Emulate (E&E). Briefly, E&E learns the low-dimensional latent embedding (like summary statistics) and a fast emulator in the latent space. Their method successfully ignores the redundant parameters and also is able to work with the Lorenz 96 system. E%E was better in performance against the NRE-C and NPE-C in terms of maximum mean discrepancy metric.

Zhang and Cao[42] closely examined a contrastive learning[20] method SimCLR on their preprint. They both theoretically worked on SimCLR and also empirically with a toy image dataset from the MNIST. SimCLR was quite beneficial in reducing the labeling complexity.

Li and Yan[43] studied the theoretical basis of the empirical success of the diffusion models with minimal assumptions. The core idea of such models is completely generative. From the one side, a sample $X_0$ is drawn from the target distribution it is progressively transformed into a pure Gaussian noise. Conversely, $Y_T$ starts from the pure Gaussian noise, and step-by-step it is converted to the target distribution at $Y_0$. X and Y samples' distributions are kept close when their subscript is equal to each other. They succeeded at providing a framework with minimal assumptions compared to other similar works in literature.

## Optimization

Liang, Neufeld, and Zhang[44] made a convergence analysis on the stochastic gradient Hamiltonian[21] Monte Carlo (SGHMC) algorithm[22] with discrete stochastic gradients. They applied the SGHMC to quantile estimation for the Gaussian, logistic, and Gumbel distribution, then, for regularization of the ReLU networks in solving optimization problems. They found it to be superior to SGLD, TUSLA, ADAM, AMSGrad, and RMSProp optimizers.

Bender and Thuan[45] worked on the reinforcement learning's modeling of exploration. They formulated a new stochastic differential equation according to the sampling on a discrete-time grid, utilized random measures in addition to the Brownian motion and Poisson random measures, and proved a limit theorem where the sampling grid mesh size approached zero. They stated that this grid-sampling limit SDE can replace exploratory SDE.

Hwang and Lim[46] improved a Physics-informed Neural Network in solving partial differential equations with their Dual Cone Gradient Descent optimizer. Their numerical experiments with three partial differential equations (PDEs) showed that this approach outperforms all other common optimizers used in the study.

Scheidt et al.[47] crafted a way to solve a False Discovery Rate-Controlled Sparse Regression problem with so high variables in a drastically less memory-requiring way, capable of being computed in a normal laptop computer. For this, they utilized memory mapping and also dummy permutating concepts to the Big T-Rex Selector.

## Probability

Wu et al.[48] worked on conditional testing with localized conformal p-values. They show example applications via a conditional outlier detection case, conditional label screening problem, and two-sample conditional distributing testing problem.

Zhang and Candès[49] introduced posterior conformal prediction (PCP)[23] method. PCP required smaller sets than the standard competitors in classification.

---

[20] First, a large amount of unlabeled data is shown to the algorithm. Then, labeling and supervising is provided as a fine-tuning.

[21] Further information on Hamiltonian and relevant concepts can be found here.

[22] The GitHub repository for the SGHMC is here.

[23] The GitHub repository to implement this method is here.

# References

[1] Thiem Hoang. Photoemission of spin-polarized electrons from aligned grains and chiral symmetry breaking, 2024, 2312.15934.

[2] Cheng-Han Hsieh, Héctor G. Arce, María José Maureira, Jaime E. Pineda, Dominique Segura-Cox, Diego Mardones, Michael M. Dunham, and Aiswarya Arun. The alma legacy survey of class 0/i disks in corona australis, aquila, chamaeleon, ophiuchus north, ophiuchus, serpens (campos). i. evolution of protostellar disk radii, 2024, 2404.02809.

[3] A S Voloshina, A D Lavrukhina, M V Pruzhinskaya, K L Malanchev, E E O Ishida, V V Krushinsky, P D Aleo, E Gangler, M V Kornilov, V S Korolev, E Russeil, T A Semenikhin, S Sreejith, and A A Volnova. Snad catalogue of m-dwarf flares from the zwicky transient facility. *Monthly Notices of the Royal Astronomical Society*, 533(4):4309–4323, August 2024.

[4] Hunter Brooks, Dan Caselden, J. Davy Kirkpatrick, Yadukrishna Raghu, Charles Elachi, Jake Grigorian, Asa Trek, Andrew Washburn, Hiro Higashimura, Aaron Meisner, Adam Schneider, Jacqueline Faherty, Federico Marocco, Christopher Gelino, Jonathan Gagné, Thomas Bickle, Shih yun Tang, Austin Rothermich, Adam Burgasser, Marc J. Kuchner, Paul Beaulieu, John Bell, Guillaume Colin, Giovanni Colombo, Alexandru Dereveanco, Deiby Flores, Konstantin Glebov, Leopold Gramaize, Les Hamlet, Ken Hinckley, Martin Kabatnik, Frank Kiwy, David Martin, Raul Palma, William Pendrill, Lizzeth Ruiz, John Sanchez, Arttu Sainio, JÖrg SchÜmann, Manfred Schonau, Christopher Tanner, Nikolaj Stevnbak Andersen, Andrés Stenner, Melina Thévenot, Vinod Thakur, Nikita Voloshin, and Zbigniew Wedracki. Discovery of 118 new ultra-cool dwarf candidates using machine learning techniques, 2024, 2408.14447.

[5] Shuai Liu, Huigang Wei, Jianrong Shi, Wenxian Li, Henggeng Han, Jifeng Liu, and Shangbin Yang. Chromospheric modeling of the active m3v star g 80-21 with rh1.5d, 2024, 2409.16028.

[6] Zhijun Tu, Shu Wang, and Jifeng Liu. Physical parameters and properties of 20 cold brown dwarfs in jwst, 2024, 2409.19191.

[7] C. Viscasillas Vázquez, E. Solano, A. Ulla, M. Ambrosch, M. A. Álvarez, M. Manteiga, L. Magrini, R. Santoveña-Gómez, C. Dafonte, E. Pérez-Fernández, A. Aller, A. Drazdauskas, Š. Mikolaitis, and C. Rodrigo. Advanced classification of hot subdwarf binaries using artificial intelligence techniques and gaia dr3 data, 2024, 2409.17783.

[8] Zhecheng Hu, Wei Zhu, Fei Dai, Ping Chen, Yang Huang, Min Fang, and Richard S. Post. An eccentric binary with a misaligned circumbinary disk, 2024, 2409.18296.

[9] Jamie Williams, Boris Gaensicke, Andrew Swan, Mairi O'Brien, Paula Izquierdo, Anna-Maria Cutolo, and Tim Cunningham. Pewdd: A database of white dwarfs enriched by exo-planetary material, 2024, 2409.16046.

[10] I. O. Eya, E. U. Iyida, O. Okike, R. E. Ugwoke, F. M. Menteso, C. J. Ugwu, P. Simpemba, J. Simfukwe, D. Silungwe S. P. Phiri, G. F Abbey, J. A. Alhassan, and A. E. Chukwude. On the simultaneity of forbush decreases: the simultaneous effects of interplanetary parameters and geomagnetic activity indices, 2024, 2409.19612.

[11] Z. Li, H. Zou, J. Liu, J. Ma, X. Zhao, X. Li, Z. Tu, B. Zhang, R. Wang, S. Wang, and Marco Xue. A multi-station meteor monitoring (m$^3$) system. i. design and testing, 2024, 2409.19503.

[12] James Sanders, October Owen, Brian Tran, Jeffrey Mosqueira, and Michael Marty. Coupling online size exclusion chromatography with charge detection-mass spectrometry using hadamard transform multiplexing. 2024.

[13] Oscar F Laurent, Maria Letizia De Marco, Marco Faustini, Cédric Boissière, Claude Poleunis, Arnaud Delcorte, and Damien P Debecker. High entropy alloys: assessing atomic-scale mixing and surface passivation with time-of-flight secondary ion mass spectrometry. 2024.

[14] Rasmus Bro and Sijmen De Jong. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 11(5):393–401, 1997.

[15] Lisa Hua, Marcello Solomon, Deanna D'Alessandro, and William Donald. Metal-organic frameworks for the trace multiplexed adsorption and quantitation of 50 per-and polyfluoroalkyl substances. 2024.

[16] Peter K Weber, Marc Debliqui, Céline Defouilloy, Xavier Mayali, Ming-Chang Liu,

Rachel Hestrin, Jennifer Pett-Ridge, Rhona Stuart, Megan Morris, Christina Ramon, et al. The nanosims-hr: the next generation of high spatial resolution dynamic sims. 2024.

[17] Stephan Freitag, Maximilian Anlanger, Maximilian Lippl, Klemens Mechtler, Elisabeth Reiter, Heinrich Grausgruber, and Rudolf Krska. Simplifying wheat quality assessment: Using near-infrared spectroscopy and analysis of variance simultaneous component analysis to study regional and annual effects. 2024.

[18] Age K Smilde, Jeroen J Jansen, Huub CJ Hoefsloot, Robert-Jan AN Lamers, Jan Van Der Greef, and Marieke E Timmerman. Anova-simultaneous component analysis (asca): a new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13):3043–3048, 2005.

[19] Mehdi Soussi-Therond, David Guarin, Aiky Razanahoera, Yongmin Zhang, Mathieu Baudin, Emeric Miclet, Nicolas Giraud, and Daniel Abergel. Simultaneous observation of the anomerization and reaction rates of enzymatic dehydrogenation of glucose-6-phosphate. 2024.

[20] Peikai Zhang, David Williams, Logan Stephens, Robert Helps, Shamini Pushparajah, Jadranka Travas-Sejdic, and Marion Wood. Microfluidic sensors for the detection of motile plant zoospores. 2024.

[21] Luigi Russo, Francesco Mauro, Alessandro Sebastianelli, Paolo Gamba, and Silvia Liberata Ullo. Sen12-water: A new dataset for hydrological applications and its benchmarking, 2024, 2409.17087.

[22] Hanbo Bi, Yingchao Feng, Yongqiang Mao, Jianning Pei, Wenhui Diao, Hongqi Wang, and Xian Sun. Agmtr: Agent mining transformer for few-shot segmentation in remote sensing, 2024, 2409.17453.

[23] Lucia Gordon, Nikhil Behari, Samuel Collier, Elizabeth Bondi-Kelly, Jackson A. Killian, Catherine Ressijac, Peter Boucher, Andrew Davies, and Milind Tambe. Find rhinos without finding rhinos: Active learning with multimodal imagery of south african rhino habitats. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI-2023, page 5977–5985. International Joint Conferences on Artificial Intelligence Organization, August 2023.

[24] Runze Zhang, Abhi Nayak, Derek Houtz, Adam Watts, Elahe Soltanaghai, and Mohamad Alipour. Sub-meter remote sensing of soil moisture using portable l-band microwave radiometer, 2024, 2409.17024.

[25] Rita Rocio Guzman-Lopez, Luis Huamanchumo, Kevin Fernandez, Oscar Cutipa-Luque, Yhon Tiahuallpa, and Helder Rojas. Sparsity, regularization and causality in agricultural yield: The case of paddy rice in peru, 2024, 2409.17298.

[26] Jiaqi Zhao, Zeyu Ding, Yong Zhou, Hancheng Zhu, Wen-Liang Du, Rui Yao, and Abdulmotaleb El Saddik. Orientedformer: An end-to-end transformer-based oriented object detector in remote sensing images, 2024, 2409.19648.

[27] Adedeji Adebukola Adelodun, Mohsen Rezaei, Siamak Ardkapan, Nan Normann Jakobsen, and Matthew Johnson. Impact of sensor placement on indoor air quality monitoring: A comparative analysis. 2024.

[28] Rachel L London, Juliane Glüge, and Martin Scheringer. Multi-criteria decision analysis for assessments of chemical alternatives (mcda-aca). 2024.

[29] Audrey Laberge-Carignan, Florence Mercier, Dominic Larivière, and Raoul-Marie Couture. Influence of temperature on selenium mobility under contrasting redox conditions: a sediment flow-through reactor experiment. 2024.

[30] Weiru Fan, Xiaobin Tang, Xingqi Xu, Huizhu Hu, Vladislav V. Yakovlev, Shi-Yao Zhu, Da-Wei Wang, and Delong Zhang. Seeing the invisible through speckle images, 2024, 2409.18815.

[31] Ronald Katende. A novel framework for analyzing structural transformation in data-constrained economies using bayesian modeling and machine learning, 2024, 2409.16738.

[32] Lauren Elborough, Duncan Taylor, and Melissa Humphries. A novel application of shapley values for large multidimensional time-series data: Applying explainable ai to a dna profile classification neural network, 2024, 2409.18156.

[33] Daan Caljon, Jeff Vercauteren, Simon De Vos, Wouter Verbeke, and Jente Van Belle. Using dynamic loss weighting to boost improvements in forecast stability, 2024, 2409.18267.

[34] Charles Marx, Volodymyr Kuleshov, and Stefano Ermon. Calibrated probabilistic forecasts for arbitrary sequences, 2024, 2409.19157.

[35] Dalin Qin, Yehui Li, Weiqi Chen, Zhaoyang Zhu, Qingsong Wen, Liang Sun, Pierre Pinson, and Yi Wang. Evolving multi-scale normalization for time series forecasting under distribution shifts, 2024, 2409.19718.

[36] Lê-Nguyên Hoang. The poison of dimensionality, 2024, 2409.17328.

[37] Han Wang and Yixuan Li. Bridging ood detection and generalization: A graph-theoretic view, 2024, 2409.18205.

[38] Richard D. Paul, Alessio Quercia, Vincent Fortuin, Katharina Nöh, and Hanno Scharr. Parameter-efficient bayesian neural networks for uncertainty-aware depth estimation, 2024, 2409.17085.

[39] Théo Bourdais and Houman Owhadi. Model aggregation: minimizing empirical variance outperforms minimizing empirical error, 2024, 2409.17267.

[40] Yu Wang, Yuxuan Yin, and Peng Li. Towards the mitigation of confirmation bias in semi-supervised learning: a debiased training perspective, 2024, 2409.18316.

[41] Ruoxi Jiang, Peter Y. Lu, and Rebecca Willett. Embed and emulate: Contrastive representations for simulation-based inference, 2024, 2409.18402.

[42] Han Zhang and Yuan Cao. Understanding the benefits of simclr pre-training in two-layer convolutional neural networks, 2024, 2409.18685.

[43] Gen Li and Yuling Yan. $o(d/t)$ convergence theory for diffusion probabilistic models under minimal assumptions, 2024, 2409.18959.

[44] Luxu Liang, Ariel Neufeld, and Ying Zhang. Non-asymptotic convergence analysis of the stochastic gradient hamiltonian monte carlo algorithm with discontinuous stochastic gradient with applications to training of relu neural networks, 2024, 2409.17107.

[45] Christian Bender and Nguyen Tran Thuan. A random measure approach to reinforcement learning in continuous time, 2024, 2409.17200.

[46] Youngsik Hwang and Dong-Young Lim. Dual cone gradient descent for training physics-informed neural networks, 2024, 2409.18426.

[47] Fabian Scheidt, Jasin Machkour, and Michael Muma. Solving fdr-controlled sparse regression problems with five million variables on a laptop. In *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAM-SAP)*, page 116–120. IEEE, December 2023.

[48] Xiaoyang Wu, Lin Lu, Zhaojun Wang, and Changliang Zou. Conditional testing based on localized conformal p-values, 2024, 2409.16829.

[49] Yao Zhang and Emmanuel J. Candès. Posterior conformal prediction, 2024, 2409.19712.