Explainable and Privacy-Preserving Machine Learning via Domain-Aware Symbolic Regression

Kei Sen Fong FONGKEISEN@U.NUS.EDU
Department of Electrical and Computer Engineering, National University of Singapore

Mehul Motani MOTANI@NUS.EDU.SG
Department of Electrical and Computer Engineering, Institute of Data Science, N.1 Institute for Health, Institute for Digital Medicine (WisDM), National University of Singapore

Abstract

Explainability and privacy are the top concerns in machine learning (ML) for medical applications. In this paper, we propose a novel method, Domain-Aware Symbolic Regression with Homomorphic Encryption (DASR-HE), that addresses both concerns simultaneously by: (i) producing domain-aware, intuitive and explainable models that do not require the end-user to possess ML expertise and (ii) training only on securely encrypted data without access to actual data values or model parameters. DASR-HE is based on Symbolic Regression (SR), which is a first-class ML approach that produces simple and concise equations for regression, requiring no ML expertise to interpret. In our work, we improve the performance of SR algorithms by using existing domain-specific medical equations to augment the search space of equations, decreasing the search complexity and producing equations that are similar in structure to those used in practice. To preserve the privacy of the medical data, we enable our algorithm to learn on data that is homomorphically encrypted (HE), meaning that arithmetic operations can be done in the encrypted space. This makes HE suitable for machine learning algorithms to learn models without access to the actual data values or model parameters. We evaluate DASR-HE on three medical tasks, namely predicting glomerular filtration rate, endotracheal tube (ETT) internal diameter and ETT depth and find that DASR-HE outperforms existing medical equations, other SR ML algorithms and other explainable ML algorithms.

Data and Code Availability In our work, we use publicly available datasets on glomerular filtration rate (GFR) measured on a population of Congolese adults (Bukabau et al., 2018) and endotracheal tubes (ETT) internal diameter and depth measured on a population of pediatric surgical patients (Kim et al., 2023). Our code is made available in the supplemental materials at: https://github.com/kentridgeai/DASR.

1. Introduction

Explainability and privacy are the top concerns in machine learning (ML) for medical applications. Even when ML models trained on medical data achieve breakthroughs in prediction performance, there is still hesitance to deploy these models due to limited explainability of the model, and concerns on the privacy of the medical data (Ahmed et al., 2023; Khan et al., 2023). In this paper, we introduce improvements to a unique ML algorithm, Symbolic Regression (SR), which finds simple explainable equations that function as predictors. The approach SR takes is orthogonal to popular black-box ML since the models are in the form of white-box equations.

Explainability: SR algorithms are ML algorithms that innately produce explainable models - in the form of concise equations that can be analyzed with ease for both regression and classification tasks (Koza, 1992; Fong and Motani, 2024). This makes SR algorithms well-suited for fields where the cost of making wrong decisions is high, such as healthcare (Christensen et al., 2022; Wilstrup and Cave, 2022). However, the equations discovered by traditional SR algorithms are independent of the application's field, producing equations which may not be suitable or widely accepted in the specific domain. To tackle this problem in Physics, researchers have seeded the SR algorithm with knowledge of existing Physics equations (Udrescu and Tegmark, 2020; Fong et al., 2023). Taking inspiration from these

works, we introduce Domain-Aware Symbolic Regression (DASR) for medical applications. DASR is a novel method which takes existing medical equations as prior knowledge and evolves them. Thus, the equations discovered by DASR have structures that are similar to existing medical equations, further increasing explainability. We also find that DASR has better prediction performance than existing SR.

Privacy: Healthcare organizations are required to keep patients' protected health information (PHI) safe. The Health Insurance Portability and Accountability Act (HIPAA) requires that all electronic PHI be encrypted when created, stored or transmitted. This introduces significant overhead when applying machine learning to healthcare data, especially when the data scientists are from an external organization or when engaging third-party computing services. Homomorphic encryption (HE) is an ideal solution to this problem, in which data is encrypted in a way that allows arithmetic operations without decryption (Munjal and Bhatia, 2023). This way, the true values of the PHI are not made known to service providers, which reduces the need for traditional security measures such as data censorship, which are costly and may degrade the performance of learning algorithms. The Cheon, Kim, Kim and Song (CKKS) scheme (Cheon et al., 2017) is the state-of-the-art algorithm for HE, and is especially suitable for ML since CKKS allows for arithmetic on floating point data (unlike most alternative HE schemes). However, the arithmetic operations are restricted to addition operations and a limited number of multiplication operations, while also adding noise. This means that it is not possible to perform certain operations directly, such as the comparison operation. Thus, existing machine learning algorithms require innovative modifications to be compatible with general HE and CKKS encrypted data (Akavia et al., 2022; Xu et al., 2023; Cong et al., 2022; Kim et al., 2018). In this work, we propose the first SR algorithm which is modified to be compatible with HE encrypted data. Specifically, we modify our DASR algorithm to operate on CKKS encrypted data in an algorithm we call Domain-Aware Symbolic Regression with Homomorphic Encryption (DASR-HE).

To evaluate DASR-HE, we work on three medical tasks of predicting (i) glomerular filtration rate (GFR), (ii) endotracheal tube (ETT) internal diameter and (iii) ETT depth. GFR is a key indicator of kidney health and is an important determinant in certain diagnoses, such as Chronic Kidney Disease (CKD) (Levey and Inker, 2016). Measuring GFR directly is expensive and time-consuming since it involves the plasma or urinary clearance of exogenous filtration markers (e.g. inulin and iohexol). Thus, it is of great interest to predict GFR from other biomarkers (Inker et al., 2021; Wang et al., 2022; Woillard et al., 2021). ETT internal diameter and ETT depth estimations are critical to reduce complications during intubation. Improper estimates used for intubation increase the risk of airway injury, prolonged apnea, pneumothorax and atelectasis (Kim et al., 2023). In pediatric patients, X-rays are not readily available, thus, it has been of great interest to predict ETT internal diameter and depth from demographic data (Zhuang et al., 2023; Topjian et al., 2020; Eipe et al., 2009; Shih et al., 2008; Cole, 1957).

The main **contributions** of this paper are:

1. We propose a novel SR algorithm, DASR, which uses known medical equations to augment the search space of equations, decreasing the search complexity and producing equations that are similar in structure to those used in the medical domain. This improves the prediction performance and explainability of the discovered equation.
2. We introduce the first SR algorithm that is modified to be compatible with HE encrypted data, which we term DASR-HE (built upon DASR). In particular, we demonstrate the effectiveness of DASR-HE on CKKS encrypted data. We also show that DASR-HE is competitive with DASR even with the added noise and constraints from CKKS.
3. We evaluate DASR and DASR-HE on 3 medical applications and show that they outperform existing medical equations, other SR ML algorithms and other explainable ML algorithms on a variety of prediction and complexity metrics.

2. Background & Related Work

2.1. Symbolic Regression

Symbolic Regression Benchmarks. SR algorithms typically use genetic programming (GP) to search through a large variety of possible equations (Koza, 1992; Fong et al., 2023; Schmidt and Lipson, 2009; Fong and Motani, 2024). Most state-of-the-art (SOTA) SR algorithms still rely on GP as the core of their algorithms (Petersen et al., 2019; Mundhenk et al., 2021). GP-based SR works by starting with an initial random population of equations, evaluating them, and modifying these equations (via prede-

fined evolutionary operations such as crossover and mutation) based on their evaluation scores. In this work, we choose deep symbolic regression (DSR) (Petersen et al., 2019), neural-guided genetic programming (NGGP) (Mundhenk et al., 2021) and DistilSR (Fong and Motani, 2023) as our choice of SOTA SR methods for comparison based on 2 recent SR benchmarks in terms of prediction and complexity by La Cava et al. (2021) and Kamienny et al. (2023). DistilSR is most similar to our approach in the sense that both explore a search space of K-Expressions. DistilSR can be said to be DASR without existing medical equations.

2.2. Existing Medical Equations

Existing GFR Equations. To evaluate the effectiveness of DASR and DASR-HE on GFR prediction, we also selected a set of benchmark clinical equations. We chose all relevant GFR equations available on MDCalc, a well-known medical reference for clinical equations (Elovic and Pourmand, 2019; Soleimanpour and Bann, 2022), used by millions of medical professionals globally (over 200 countries), inclusive of more than 65% of US physicians. The equations are made available in Table 1, consisting of MDRD (Levey et al., 2006), Schwartz equation (Schwartz et al., 2009), CKD-EPI Creatinine (Levey et al., 2009), CKD-EPI Cystatin and CKD-EPI Creatinine-Cystain C (Inker et al., 2012, 2021). When ethnic factor correction is available in the equation, we present both versions: (i) with ethnic factor (WEF) (ii) no ethnic factor (NEF).

Existing ETT Equations. To evaluate the effectiveness of DASR and DASR-HE on ETT internal diameter and ETT depth prediction, we also selected a set of benchmark clinical equations. For ETT internal diameter, we base our selection on a recent comparative study (Subramani et al., 2023). The equations are made available in Table 2, consisting of age-based formula (ABF) (Cole, 1957), height-based formula (HBF) (Shih et al., 2008) and weight-based formula (WBF) (Eipe et al., 2009). For ETT depth, the equations are made available in Table 3, consisting of pediatric advanced life support formula (PALS) (Topjian et al., 2020) and height-based formula 2 (HBF-2) (Zhuang et al., 2023).

2.3. Homomorphic Encryption

CKKS Encyption Scheme. The Cheon, Kim, Kim and Song (CKKS) scheme (Cheon et al., 2017) is the state-of-the-art algorithm for HE. The CKKS scheme is both additively and multiplicatively homomorphic, meaning that given messages m_1, m_2, and their corresponding CKKS-encrypted version, $\mathcal{E}(m_1), \mathcal{E}(m_2)$, the following 2 properties generally holds: $\mathcal{E}(m_1) + \mathcal{E}(m_2) = \mathcal{E}(m_1 + m_2)$ and $\mathcal{E}(m_1) \times \mathcal{E}(m_2) = \mathcal{E}(m_1 \times m_2)$. This means that addition and multiplication can be done on encrypted data without requiring decryption. CKKS is the best suited HE for ML because it supports floating point arithmetic, unlike other SOTA HE schemes like BGV (Yagisawa, 2015) and BFV (Fan and Vercauteren, 2012) which only supports integer arithmetic operations. Thus, CKKS is the most popular choice for HE for ML algorithms (Akavia et al., 2022; Xu et al., 2023; Cong et al., 2022; Kim et al., 2018). However, in CKKS, there is a limit to the number of mathematical operations that can be performed on an encrypted data until the noise becomes too large, especially with the multiplication operator. We can increase the limit to the number of multiplications, but this comes at the cost of longer compute time, or lower security level (Albrecht et al., 2021). Thus, ML algorithms on CKKS-encrypted data need to manage the operations carefully, which we do in our work.

Other Operations in HE. Since HE only supports addition and multiplication, many non-polynomial operations require innovative methods or approximations. For example, comparison operators (i.e., $comp(a, b)$, which outputs 1 if $a > b$ and 0 if $a < b$) are a crucial component in many ML algorithms, where it is often required in the training and inference stages. In HE, it is challenging to generate a comparison function with only addition and multiplication. One solution is to send the entire operation and operands to the secret-key holder and request the evaluation of the whole operation, which will then be encrypted and sent back. Another solution is to use composite polynomials to approximate the comparison operators to a high degree of precision (Cheon et al., 2020). Similar solutions exists for other functions such as division and exponentiation (Babenko and Golimblevskaia, 2021; Prantl et al., 2023).

3. Methodology

DASR Details. In traditional SR algorithms, the equations found are usually drastically different from existing medical equations, which reduces explainability, and does not exploit existing domain knowledge. To address this, we introduce DASR, which

Table 1: **Existing Equations for Predicting GFR.** 10 commonly used GFR equations from MDCalc. When ethnic factor correction is available, we present both versions: (i) with ethnic factor (WEF) (ii) no ethnic factor (NEF). The features used in the equations are {*age* in years, *gender*, serum creatinine in mg/dL (SCR), serum cystatin C in mg/L ($SCYS$), *height* in cm}.

Name of Equation	Condition	Simplified Equation
MDRD NEF (Levey et al., 2006)	female	$129.85 \times SCR^{-1.154} \times age^{-0.203}$
	male	$175 \times SCR^{-1.154} \times age^{-0.203}$
MDRD WEF (Levey et al., 2006)	female	$157.37 \times SCR^{-1.154} \times age^{-0.203}$
	male	$212.1 \times SCR^{-1.154} \times age^{-0.203}$
Schwartz Equation (Schwartz et al., 2009)	all	$0.413 \times height/SCR$
CKD-EPI Creatinine NEF (Levey et al., 2009)	female, $SCR \leq 0.7$	$128.06 \times SCR^{-0.329} \times 0.993^{age}$
	female, $SCR > 0.7$	$93.559 \times SCR^{-1.209} \times 0.993^{age}$
	male, $SCR \leq 0.9$	$135.02 \times SCR^{-0.411} \times 0.993^{age}$
	male, $SCR > 0.9$	$124.14 \times SCR^{-1.209} \times 0.993^{age}$
CKD-EPI Creatinine WEF (Levey et al., 2009)	female, $SCR \leq 0.7$	$148.42 \times SCR^{-0.329} \times 0.993^{age}$
	female, $SCR > 0.7$	$108.43 \times SCR^{-1.209} \times 0.993^{age}$
	male, $SCR \leq 0.9$	$156.49 \times SCR^{-0.411} \times 0.993^{age}$
	male, $SCR > 0.9$	$143.87 \times SCR^{-1.209} \times 0.993^{age}$
CKD-EPI Cystatin C (Inker et al., 2012)	female, $SCYS \leq 0.8$	$110.89 \times SCYS^{-0.499} \times 0.996^{age}$
	female, $SCYS > 0.8$	$92.166 \times SCYS^{-1.328} \times 0.996^{age}$
	male, $SCYS \leq 0.8$	$118.98 \times SCYS^{-0.499} \times 0.996^{age}$
	male, $SCYS > 0.8$	$98.89 \times SCYS^{-1.328} \times 0.996^{age}$
CKD-EPI Creatinine-Cystatin C NEF (Inker et al., 2012)	female, $SCYS \leq 0.8$, $SCR \leq 0.7$	$109.44 \times SCR^{-0.248} \times SCYS^{-0.375} \times 0.995^{age}$
	female, $SCYS \leq 0.8$, $SCR > 0.7$	$96.495 \times SCR^{-0.601} \times SCYS^{-0.375} \times 0.995^{age}$
	female, $SCYS > 0.8$, $SCR \leq 0.7$	$101.53 \times SCR^{-0.248} \times SCYS^{-0.711} \times 0.995^{age}$
	female, $SCYS > 0.8$, $SCR > 0.7$	$89.524 \times SCR^{-0.601} \times SCYS^{-0.711} \times 0.995^{age}$
	male, $SCYS \leq 0.8$, $SCR \leq 0.9$	$121.48 \times SCR^{-0.207} \times SCYS^{-0.375} \times 0.995^{age}$
	male, $SCYS \leq 0.8$, $SCR > 0.9$	$116.54 \times SCR^{-0.601} \times SCYS^{-0.375} \times 0.995^{age}$
	male, $SCYS > 0.8$, $SCR \leq 0.9$	$112.71 \times SCR^{-0.207} \times SCYS^{-0.711} \times 0.995^{age}$
	male, $SCYS > 0.8$, $SCR > 0.9$	$108.12 \times SCR^{-0.601} \times SCYS^{-0.711} \times 0.995^{age}$
CKD-EPI Creatinine-Cystatin C WEF (Inker et al., 2012)	female, $SCYS \leq 0.8$, $SCR \leq 0.7$	$118.19 \times SCR^{-0.248} \times SCYS^{-0.375} \times 0.995^{age}$
	female, $SCYS \leq 0.8$, $SCR > 0.7$	$104.21 \times SCR^{-0.601} \times SCYS^{-0.375} \times 0.995^{age}$
	female, $SCYS > 0.8$, $SCR \leq 0.7$	$109.66 \times SCR^{-0.248} \times SCYS^{-0.711} \times 0.995^{age}$
	female, $SCYS > 0.8$, $SCR > 0.7$	$96.686 \times SCR^{-0.601} \times SCYS^{-0.711} \times 0.995^{age}$
	male, $SCYS \leq 0.8$, $SCR \leq 0.9$	$131.2 \times SCR^{-0.207} \times SCYS^{-0.375} \times 0.995^{age}$
	male, $SCYS \leq 0.8$, $SCR > 0.9$	$125.86 \times SCR^{-0.601} \times SCYS^{-0.375} \times 0.995^{age}$
	male, $SCYS > 0.8$, $SCR \leq 0.9$	$121.72 \times SCR^{-0.207} \times SCYS^{-0.711} \times 0.995^{age}$
	male, $SCYS > 0.8$, $SCR > 0.9$	$116.77 \times SCR^{-0.601} \times SCYS^{-0.711} \times 0.995^{age}$
CKD-EPI Creatinine (Inker et al., 2021)	female, $SCR \leq 0.7$	$131.86 \times SCR^{-0.241} \times 0.9938^{age}$
	female, $SCR > 0.7$	$93.667 \times SCR^{-1.2} \times 0.9938^{age}$
	male, $SCR \leq 0.9$	$137.55 \times SCR^{-0.302} \times 0.9938^{age}$
	male, $SCR > 0.9$	$125.13 \times SCR^{-1.2} \times 0.9938^{age}$
CKD-EPI Creatinine-Cystatin C (Inker et al., 2021)	female, $SCYS \leq 0.8$, $SCR \leq 0.7$	$111.87 \times SCR^{-0.219} \times SCYS^{-0.323} \times 0.9961^{age}$
	female, $SCYS \leq 0.8$, $SCR > 0.7$	$99.63 \times SCR^{-0.544} \times SCYS^{-0.323} \times 0.9961^{age}$
	female, $SCYS > 0.8$, $SCR \leq 0.7$	$101.07 \times SCR^{-0.219} \times SCYS^{-0.778} \times 0.9961^{age}$
	female, $SCYS > 0.8$, $SCR > 0.7$	$90.011 \times SCR^{-0.544} \times SCYS^{-0.778} \times 0.9961^{age}$
	male, $SCYS \leq 0.8$, $SCR \leq 0.9$	$123.72 \times SCR^{-0.144} \times SCYS^{-0.323} \times 0.9961^{age}$
	male, $SCYS \leq 0.8$, $SCR > 0.9$	$118.61 \times SCR^{-0.544} \times SCYS^{-0.323} \times 0.9961^{age}$
	male, $SCYS > 0.8$, $SCR \leq 0.9$	$111.77 \times SCR^{-0.144} \times SCYS^{-0.778} \times 0.9961^{age}$
	male, $SCYS > 0.8$, $SCR > 0.9$	$107.16 \times SCR^{-0.544} \times SCYS^{-0.778} \times 0.9961^{age}$

Table 2: **Existing Equations for Predicting ETT Internal Diameter.** The features are {*age* in years, *height* in cm, *weight* in kg}.

Name of Equation	Simplified Equation
ABF (Cole, 1957)	$age/4 + 4$
HBF (Shih et al., 2008)	$height/30 + 2$
WBF (Eipe et al., 2009)	$weight/10 + 3.5$

Table 3: **Existing Equations for Predicting ETT Depth.** The features used are {*age* in years, *height* in cm}.

Name of Equation	Simplified Equation
PALS (Topjian et al., 2020)	$age/2 + 12$
HBF-2 (Zhuang et al., 2023)	$height \times 0.1 + 4$

takes existing medical equations as prior knowledge and evolves them as outlined in Algorithm 1. In DASR, we utilize K-Expressions from Gene Expression Programming (Ferreira, 2002) in order to have an easily manipulated representation of an equation (see Appendix B for details). K-Expressions have many properties which are desirable, such as having fixed length K-Expressions represent multiple equations of varying length and also fulfilling the closure property of evolutionary algorithms without much overhead (Ferreira, 2002; Fong and Motani, 2023). In DASR, we take existing medical equations applicable for the task (see examples in Tables 1,2 & 3) and convert them to their K-Expression form. The K-Expressions are then perturbed at 2 points and converted back into standard infix equations. The numerical constants of these equations are then optimized via the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) (Broyden, 1970) with respect to the mean-squared error (MSE).

By using existing medical equations as prior knowledge, DASR discovers equations that closely resemble the functional forms familiar to medical professionals. Another benefit is that DASR has a drastically smaller search space compared to traditional SR algorithms, but yet has better prediction performance as shown later in the results section. For our algorithms, we set the functions in the primitive symbols set to $\{+, -, \times, /, \wedge\}$. The hyperparameters are tuned based on mean squared error evaluated on validation data.

Algorithm 1: *DASR* Pseudo Code

Input: *existing_equations*, *primitive_symbols_set*, $\mathbf{X}$, $\mathbf{y}$, where $\mathbf{X}$ is the features and $\mathbf{y}$ is the output
Output: *best_modified_equation*
best_score ← *null*
best_modified_equation ← *null*
for *equation* ∈ *existing_equations* **do**
  existing_K_exp ← `ConvertToKExpression`(*equation*)
  max_len ← `Length`(*existing_K_exp*)
  for $i \in \{1, 2, \cdots, max\_len\}$ **do**
    for $j \in \{i+1, i+2, \cdots, max\_len\}$ **do**
      for $\alpha, \beta \in primitive\_symbols\_set^2$ **do**
        modified_K_exp ← *existing_K_exp*
        modified_K_exp[i] ← α
        modified_K_exp[j] ← β
        modified_K_exp ← `AppendPads`(*modified_K_exp*)
        `/* Padding ensures K-Expression can be validly decoded */`
        modified_equation ← `Decode`(*modified_K_exp*)
        modified_equation ← `BFGS`(*modified_equation*, $\mathbf{X}, \mathbf{y}$)
        `/* BFGS is a method for optimizing numerical constants */`
        current_score ← `MSE`(*modified_equation*, $\mathbf{X}, \mathbf{y}$)
        if *best_score* > *current_score* **then**
          best_score ← *current_score*
          best_modified_equation ← *modified_equation*
        end
      end
    end
  end
end
return *best_modified_equation*

Algorithm 2: *DASR-HE* Constants Optimization Pseudo Code (To Replace BFGS)

Input: *modified_equation*, *step_sizes*, *step_iterations*, $\mathcal{E}(\mathbf{X})$, $\mathcal{E}(\mathbf{y})$, where $\mathcal{E}(\mathbf{X})$ is the CKKS-encrypted features and $\mathcal{E}(\mathbf{y})$ is the CKKS-encrypted output
Output: *coeff*

$n \leftarrow$ `CountCoefficients`(*modified_equation*)
$coeff \leftarrow$ `RandomList`(n)
$cost \leftarrow$ `MSE`(*modified_equation*, $coeff, \mathcal{E}(\mathbf{X}), \mathcal{E}(\mathbf{y})$)
for $step, iteration \in step\_sizes,\ step\_iterations$ **do**
  for $i \in \{1, 2, \cdots, iteration\}$ **do**
    for $idx \in \{1, 2, \cdots, n\}$ **do**
      $new\_coeff \leftarrow coeff$
      $new\_coeff[idx] \leftarrow new\_coeff[idx] + step$
      $new\_cost \leftarrow$ `MSE`(*modified_equation*, $new\_coeff, \mathcal{E}(\mathbf{X}), \mathcal{E}(\mathbf{y})$)
      $outcome \leftarrow step \times (new\_cost \lesssim cost)$
      $coeff[idx] \leftarrow coeff[idx] + outcome$
      $cost \leftarrow$ `MSE`(*modified_equation*, $coeff, \mathcal{E}(\mathbf{X}), \mathcal{E}(\mathbf{y})$)
      $new\_coeff \leftarrow coeff$
      $new\_coeff[idx] \leftarrow new\_coeff[idx] - step$
      $new\_cost \leftarrow$ `MSE`(*modified_equation*, $new\_coeff, \mathcal{E}(\mathbf{X}), \mathcal{E}(\mathbf{y})$)
      $outcome \leftarrow step \times (new\_cost \lesssim cost)$
      $coeff[idx] \leftarrow coeff[idx] - outcome$
      $cost \leftarrow$ `MSE`(*modified_equation*, $coeff, \mathcal{E}(\mathbf{X}), \mathcal{E}(\mathbf{y})$)
    end
  end
end
return *coeff*

DASR-HE Details. To modify DASR to be compatible with CKKS-encrypted data, the steps in DASR have to be managed differently since the primitive operations are restricted to additions and multiplications, and these operations are limited. Thus, the numerical optimization step in DASR, which uses BFGS, is not well-suited since it requires expensive matrix inversions or numerical approximation of the jacobian which is not practical given that the CKKS scheme introduces noise during encoding and arithmetic operations (see Appendix A for details). Therefore, we replace BFGS with a method inspired by pattern search (Hooke and Jeeves, 1961). Our replacement for BFGS is shown in Algorithm 2, and is motivated by works which show that non-gradient based optimizers perform comparably with gradient based methods (Chiang et al., 2023), and it also mostly relies on few simple addition and multiplication steps. By using Algorithm 2 instead of BFGS, we comfortably optimize the coefficients for candidate expressions under the practical constraints of CKKS encryption. Note that while there seems to be redundant steps such as duplicating the coefficients in Algorithm 2, these are done to minimize the amount of noise added to the CKKS-encrypted coefficients.

For other operations, such as comparison, we can use an approximation via composite polynomials (Cheon et al., 2020), or request an evaluation from the secret-key holder (Sarpatwar et al., 2020), which we do in this work. For the CKKS encryption, we set the coefficient modulus bit sizes to $[60, 40, 40, 40, 40, 40, 40, 40, 60]$ to generate 9 prime numbers for sufficient multiplicative depth. To maintain the security level as shown in Table 1 of the homomorphic encryption standard (Albrecht et al., 2021), it is necessary to set a high polynomial modulus degree of 16384 to accommodate to the coefficient modulus bit sizes. Detailed settings for DASR-HE and CKKS are provided in the code implementation of our algorithm found in Supplemental Materials.

Datasets Details. In our work, we use datasets on glomerular filtration rate (GFR) measured on a population of Congolese adults (Bukabau et al., 2018). This dataset obtained GFR by measuring plasma clearance of iohexol. Other features in the dataset include {*age* in years, *gender*, serum creatinine in mg/dL (*SCR*), serum cystatin C in mg/L (*SCYS*), *height* in cm}, which are present in equations in Table 1. We also use datasets on endotracheal tubes (ETT) internal diameter and depth measured on a population of pediatric surgical patients (Kim et al., 2023). Other features in the dataset include {*age* in years, *height* in cm, *weight* in kg}, which are present in equations in Tables 2 & 3.

Benchmark Methods. To do a thorough evaluation, we identified 3 broad categories of competing methods: (i) other SR ML, (ii) other explainable ML and (iii) existing medical equations. We chose the 3 SOTA SR algorithms identified in related works, DSR, NGGP and DistilSR. We also chose 3 ML methods that are widely regarded as explainable (Whig et al., 2023): Linear Regression (LR), Support Vector Regression (SVR) and Decision Tree Regression (DTR), based on recent medical research utilizing the models, respectively (Raynaud et al., 2023; Zhou

Table 4: New Equations (Our Contributions) for GFR Prediction

SR Method	Condition	Our Discovered Equation
DASR (Ours)	female, $SCYS \leq 0.8$, $SCR \leq 0.7$	$SCR^{-0.23201} \times (92.956 \times SCYS^{-0.44557} - 0.30801 \times age)$
	female, $SCYS \leq 0.8$, $SCR > 0.7$	$SCR^{-0.40425} \times (79.963 \times SCYS^{-0.67882} - 0.32564 \times age)$
	female, $SCYS > 0.8$, $SCR \leq 0.7$	$SCR^{0.0115} \times (86.741 \times SCYS^{-0.097012} - 0.070911 \times age)$
	female, $SCYS > 0.8$, $SCR > 0.7$	$SCR^{0.18054} \times (112.28 \times SCYS^{-0.29257} - 0.68316 \times age)$
	male, $SCYS \leq 0.8$, $SCR \leq 0.9$	$SCR^{0.65914} \times (2.0637 \times SCYS^{1.3969} + 4.4278 \times age)$
	male, $SCYS \leq 0.8$, $SCR > 0.9$	$SCR^{-11.75} \times (0.14873 \times SCYS^{-13.426} + 1.3436 \times age)$
	male, $SCYS > 0.8$, $SCR \leq 0.9$	$SCR^{0.21501} \times (114.67 \times SCYS^{-0.31784} - 0.33236 \times age)$
	male, $SCYS > 0.8$, $SCR > 0.9$	$SCR^{-0.068742} \times (108.45 \times SCYS^{-0.12809} - 0.37943 \times age)$
DASR-HE (Ours)	female, $SCR \leq 0.7$	$298.76 \times age^{-0.11564} \times 0.40507^{SCYS}$
	female, $SCR > 0.7$	$300.3 \times age^{-0.21147} \times 0.5624^{SCYS}$
	male, $SCR \leq 0.9$	$273.6 \times age^{-0.17364} \times 0.6579^{SCYS}$
	male, $SCR > 0.9$	$242.21 \times age^{-0.17131} \times 0.71114^{SCYS}$
DSR	all	$height/(SCR + SCYS^2/SCR)$
NGGP	all	$age \times height/(age \times SCYS^2 + age + 3 \times is\_female - SCYS)$
DistilSR	all	$height^{0.8677^{SCYS}}$

Table 5: New Equations for ETT Diameter

SR Method	Our Discovered Equation
DASR (Ours)	$age^{0.48872} + 3.1074$
DASR-HE (Ours)	$age^{0.49014} + 3.1036$
DSR	$14.551 \times height/(height + 184.01)$
NGGP	$14.247 \times height/(height + 174.12)$
DistilSR	$\lvert height - 34.603\rvert^{0.39193}$

Table 6: New Equations for ETT Depth

SR Method	Our Discovered Equation
DASR (Ours)	$4.9208 + 0.094789 \times height$
DASR-HE (Ours)	$0.79881 \times height^{0.62999}$
DSR	$43.233 \times height/(height + 187.06)$
NGGP	$height/(0.03647 \times height - 0.03647 \times weight + 3.7496)$
DistilSR	$(age + height)^{0.57855}$

et al., 2023; Shikha and Kasem, 2023). Finally, we also compare against existing medical equations for each of the 3 medical prediction problems as outlined in related works Tables 1,2 & 3.

Benchmark Metrics. We measure the performance of predictors via a diverse range of prediction and complexity metrics. Based on existing medical literature, we chose the following 6 prediction metrics (where y is the actual values, $\hat{y}$ is the predictions, ρ is the pearson's correlation between actual values and predictions, μ is the mean and σ is the standard deviation):

i) Root-MSE (RMSE), $\sqrt{MSE}$.
ii) Mean absolute error (MAE), $\sum_{i=1}^{N} |y_i - \hat{y}_i|/N$.
iii) Lin's concordance correlation coefficient (CCC), $2\rho\sigma_y\sigma_{\hat{y}}/((\mu_y - \mu_{\hat{y}})^2 + \sigma_y^2 + \sigma_{\hat{y}}^2)$ (Lawrence and Lin, 1989).
iv) Proportion of predictions within ±10% of actual value (P10) (Bukabau et al., 2018).
v) Proportion of predictions within ±30% of actual value (P30) (Bukabau et al., 2018).
vi) (Only for GFR) Stage accuracy, the agreement of prediction with actual values in categorizing individuals into the 5 guideline-recommended GFR stages (Stage 1: > 90, Stage 2: 60 to 89, Stage 3: 30 to 59, Stage 4: 15 to 29, Stage 5: < 15) (Inker et al., 2021).

In terms of complexity, we chose the following two metrics from SR literature: (i) Peterson's complexity, which is the sum of pre-defined scores assigned to equation tokens as detailed in (Petersen et al., 2019) and (ii) equation length, which is the sum of occurrences of operations, constants and features in the equation (La Cava et al., 2021).

Table 7: Prediction performance of our methods against benchmarks on GFR data. Best values are bolded.

	RMSE (lower is better)	MAE (lower is better)	Lin's CCC (higher is better)	P10 (higher is better)	P30 (higher is better)	Stage Accuracy (higher is better)
Our Discovered Equations						
DASR	**13.025**	**8.711**	**0.667**	**0.632**	0.948	0.755
DASR-HE	13.63	9.425	0.623	0.622	**0.949**	**0.776**
Other SR ML						
DSR	17.294	13.217	0.481	0.438	0.908	0.591
NGGP	15.893	12.086	0.464	0.469	0.938	0.622
DistilSR	15.95	12.235	0.37	0.438	0.938	0.612
Other Explainable ML						
LR	15.491	11.303	0.503	0.52	0.948	0.673
SVR	15.073	11.174	0.49	0.52	0.918	0.683
DTR	14.755	11.086	0.528	0.52	0.938	0.602
Existing Medical Equations						
MDRD NEF (2006)	24.704	16.04	0.355	0.387	0.836	0.581
MDRD WEF (2006)	32.394	20.496	0.28	0.306	0.765	0.653
Schwartz Equation (2009)	23.144	18.579	0.196	0.244	0.785	0.418
CKD-EPI Creatinine NEF (2009)	19.564	13.735	0.512	0.5	0.806	0.663
CKD-EPI Creatinine WEF (2009)	28.45	22.536	0.365	0.204	0.704	0.653
CKD-EPI Cystatin C (2012)	21.042	15.832	0.521	0.428	0.836	0.581
CKD-EPI Creatinine-Cystatin C NEF (2012)	16.58	11.825	0.591	0.5	0.908	0.663
CKD-EPI Creatinine-Cystatin C WEF (2012)	19.975	15.409	0.518	0.367	0.836	0.653
CKD-EPI Creatinine (2021)	20.083	14.416	0.493	0.397	0.806	0.673
CKD-EPI Creatinine-Cystatin C (2021)	17.726	13.278	0.545	0.418	0.867	0.663

Table 8: Prediction performance of our methods against benchmarks for ETT internal diameter data.

	RMSE (lower is better)	MAE (lower is better)	Lin's CCC (higher is better)	P10 (higher is better)	P30 (higher is better)
Our Discovered Equations					
DASR	**0.381**	**0.31**	**0.928**	0.778	**0.999**
DASR-HE	**0.381**	**0.31**	**0.928**	0.778	**0.999**
Other SR ML					
DSR	0.387	0.312	0.928	**0.784**	0.998
NGGP	0.405	0.327	0.923	0.76	0.998
DistilSR	0.393	0.314	0.926	0.772	0.997
Other Explainable ML					
LR	0.407	0.329	0.922	0.763	0.998
SVR	0.949	0.819	0.225	0.29	0.842
DTR	0.42	0.328	0.912	0.621	**0.999**
Existing Medical Equations					
ABF (1957)	0.522	0.43	0.867	0.644	0.936
HBF (2008)	0.559	0.464	0.861	0.561	0.98
WBF (2009)	0.887	0.617	0.742	0.518	0.916

4. Results and Discussion

In Tables 4, 5 & 6, we document the newly discovered equations we generated using DASR and DASR-HE and the 3 other SR ML methods (i.e., DSR, NGGP, DistilSR) for the tasks of predicting GFR, ETT internal diameter and ETT depth, respectively. The prediction performance of all methods on the three tasks in terms of the benchmark metrics are recorded in Tables 7, 8 & 9 respectively.

Do DASR and DASR-HE predict better than other approaches? In all but 1 of the 16 prediction performance metrics, the best equations were our discovered equations using either DASR or DASR-

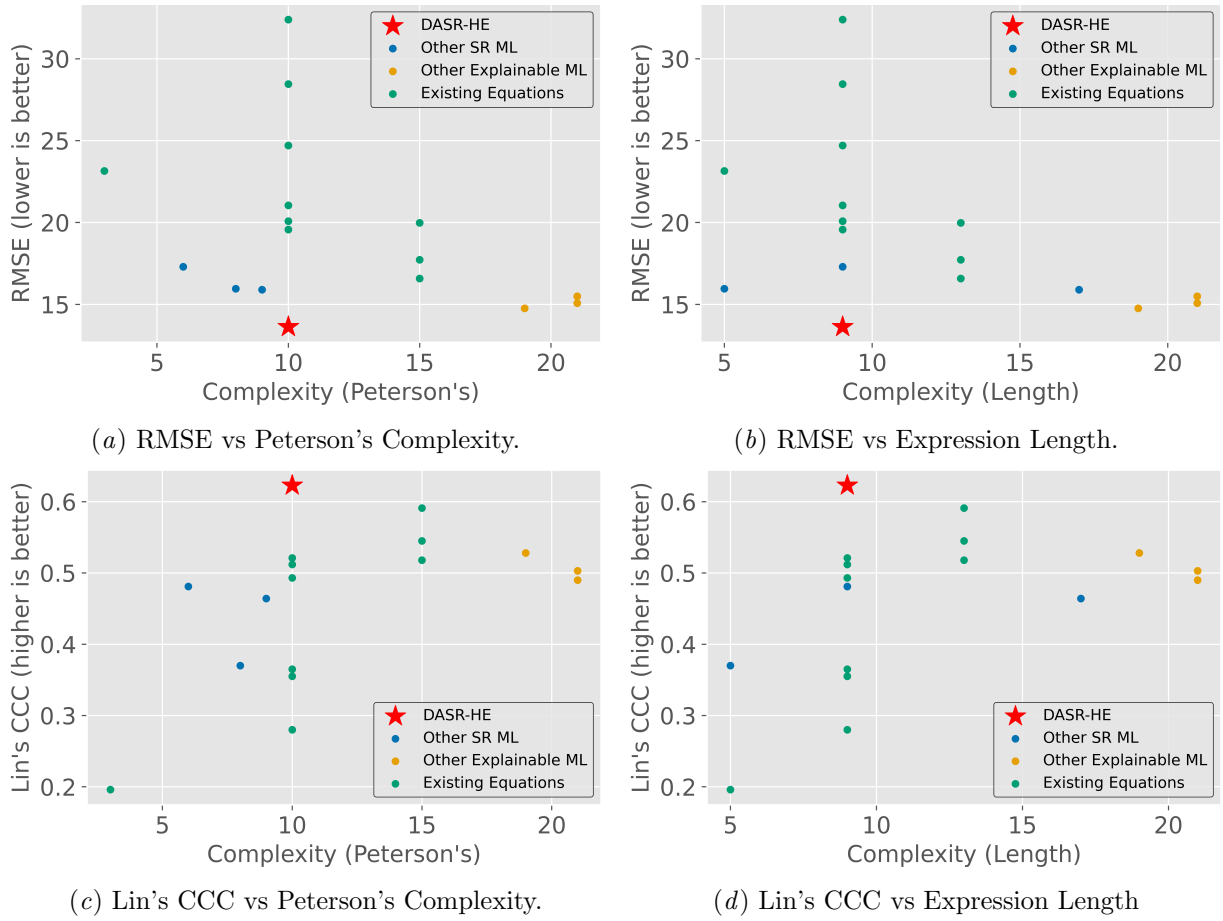
Table 9: Prediction performance of our methods against benchmarks for ETT depth data.

	RMSE (lower is better)	MAE (lower is better)	Lin's CCC (higher is better)	P10 (higher is better)	P30 (higher is better)
Our Discovered Equations					
DASR	**1.029**	**0.781**	**0.938**	**0.856**	0.997
DASR-HE	1.034	0.787	0.936	0.854	**0.998**
Other SR ML					
DSR	1.119	0.862	0.931	0.804	0.997
NGGP	1.334	0.84	0.908	0.83	0.997
DistilSR	1.03	0.788	0.937	0.848	0.997
Other Explainable ML					
LR	1.091	0.839	0.932	0.813	0.997
SVR	2.807	2.317	0.207	0.329	0.887
DTR	1.268	1.01	0.904	0.703	0.994
Existing Medical Equations					
PALS (2020)	1.504	1.194	0.825	0.667	0.967
HBF-2 (2023)	1.115	0.857	0.932	0.824	**0.998**

HE. Even in the single exception (P10 for ETT internal diameter), the performances of both DASR and DASR-HE are the next best. Notably, across Tables 7 and 8, DASR and DASR-HE outperforms better than other SR ML and other explainable ML and severely outperforms existing medical equations. In Table 9, existing medical equations performance are competitive, but DASR and DASR-HE still demonstrate the best performance among all approaches.

What if there are no existing medical equations for DASR? The DASR algorithm without existing medical equations will have to do an exhaustive search of K-Expressions, since there are no base medical equations to perturbate on. Thus, since all possible perturbations are explored, this is similar to DistilSR. While DASR searches K-Expressions 2 perturbations away from the base medical equations, DistilSR will search all possible equations, which is computational expensive (i.e., search complexity of $O(p^2l^2)$ for DASR and $O(p^l)$ for DistilSR, where p is the number of unique primitive operations and operands and l is the length of K-expressions). The poor scaling of the search complexity of DistilSR restricts the max length of expressions due to computational constraints, thus it is not viable to use DistilSR to search the space of long K-Expressions, restricting the expressivity of the discovered equation, which in turn reduces prediction performance. As seen in Tables 7, 8, 9, DASR outperforms DistilSR and DistilSR outperforms existing medical equations. Furthermore, in the workflow of discovering a medical equation, the researcher using DASR can apply simple traditional approaches of finding medical equations to get a base equation. For example, linear regression or linear SVM could be applied on the dataset. The researcher can utilize their domain knowledge to prune the insignificant weights of the linear model, do feature selection and also apply transforms to the features themselves (e.g., do $y = \sum_{i=0}^{n} a_i \mathbf{ln}(\mathbf{x})_i$ instead of $y = \sum_{i=0}^{n} a_i \mathbf{x}_i$), or apply all of the above. This will then form the base equation for DASR, in which DASR will perturb this equation to find an improved equation structure. Nonetheless, we note that medical equations are used in many specialties (see Appendix C).

Does DASR-HE sacrifice explainability to obtain high prediction performance? To show that the high prediction performance of DASR-HE is not at the expense of having low explainability (which we measure through complexity metrics), we analyze the prediction-complexity tradeoff in Figure 1 and Appendix Figures 5 & 6, for the 3 medical applications respectively. DASR-HE consistently pareto dominates all other methods. Relative to the other methods, the high prediction performance of DASR-HE is more than proportionate to its complexity. We also note that DASR-HE is able to obtain high prediction performance that even the more complex models are unable to obtain. Thus, despite not always possessing the lowest complexity, DASR-HE is the best overall approach considering its top prediction performance and relatively low complexity.



(*a*) RMSE vs Peterson's Complexity.

(*b*) RMSE vs Expression Length.

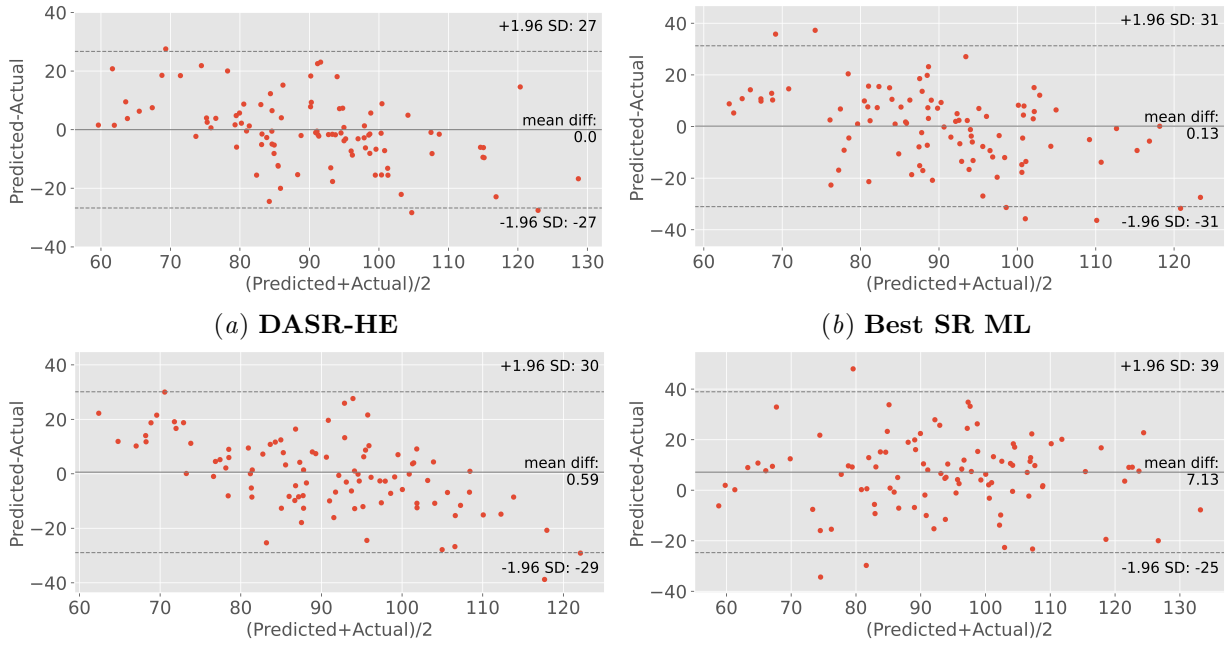
(*c*) Lin's CCC vs Peterson's Complexity.

(*d*) Lin's CCC vs Expression Length

Figure 1: Plots of various prediction metrics against various complexity metrics for GFR prediction. DASR-HE pareto dominates all other approaches, in terms of performance and complexity.

How does DASR-HE perform in terms of bias and outliers? The Bland-Altman analysis is commonly used in healthcare to evaluate the degree of agreement between a prediction and the actual value (Bukabau et al., 2018), by identifying systematic bias and influential outliers. We conducted the Bland-Altman for DASR-HE, the best SR ML, the best explainable ML and the best existing medical equation in the plots in Fig. 2. In all plots except DASR-HE, we see that the mean difference is highly positive, suggesting that other approaches tend to have overestimated predictions. Additionally, the standard deviation of the difference between predicted and actual values is the smallest in DASR-HE, which is also consistent with the high P10 and P30 scores obtained by DASR-HE. Furthermore, despite the smaller standard deviation, most of the differences lie within ± 1.96 standard deviation (95% limits of agreement).

Ablation: How much performance does DASR-HE sacrifice for the increased privacy? We should expect that DASR-HE performs worse than DASR since the purpose of encrypting the data is to preserve privacy and comes at the expense of increased noise. Note that the noise added during encryption is important for the security of the encryption and is not an unwanted removable side-effect. DASR-HE performs worse than DASR because the noise introduced by CKKS affects the MSE computation of candidate equations and changes their rankings, which is more apparent in long equations with many operations. We can observe this in Table 4, where the equations discovered by DASR-HE



(*a*) **DASR-HE** (*b*) **Best SR ML**

(*c*) **Best Explainable ML** (*d*) **Best Existing Medical Equation**

Figure 2: Bland-Altman analysis DASR-HE and the top performer among the 3 categories of SR ML, Explainable ML and Existing Medical Equations.

is shorter than that of DASR. From the prediction of ETT internal diameter and ETT depth, we can see in Tables 8 & 9 that DASR-HE performance metrics are within 1% of DASR. For prediction of GFR, DASR-HE prediction performance is slightly worse, but still performs better than all other methods, coming only second to DASR. Despite the additional noise due to CKKS-encryption and CKKS arithmetic operations, the performance of DASR-HE still remains close to that of DASR on all three applications, demonstrating the success and robustness of our modification from DASR to DASR-HE.

Limitations: Our work is based upon retrospective data and hence future work on a prospective cohort is required to provide further validation for the equations discovered by our methods.

5. Conclusion

In this paper, we propose DASR-HE, a novel SR ML algorithm that is both explainable and privacy-preserving. We first develop a base SR algorithm without encryption, DASR, which uses known medical equations to augment the search space of equations, decreasing the search complexity and producing equations that are similar in structure to those used in medical practice. These new equations show high prediction performance and explainability. Then, we introduce a first-of-its-kind SR algorithm (a modification to DASR) that is compatible with CKKS-encrypted data, which we term DASR-HE. We show that DASR-HE is competitive with DASR even with the added noise and constraints from CKKS. Finally, we evaluate DASR and DASR-HE on 3 medical applications and show that they outperform existing medical equations, other SR ML algorithms and other explainable ML algorithms on a range of prediction and complexity metrics. We hope that our work will motivate healthcare professionals to utilize DASR-HE to discover new explainable medical equations in an automated privacy-preserving manner. DASR-HE takes an orthogonal approach to common ML algorithms - by producing models (equations) that do not require intricate ML knowledge to interpret.

Institutional Review Board (IRB) This research does not require IRB approval.

Acknowledgments

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-PhD-2023-08-052T), and A*STAR, CISCO Systems (USA) Pte. Ltd and National University of Singapore under its Cisco-NUS Accelerated Digital Economy Corporate Laboratory (Award I21001E0002).

References

Molla Imaduddin Ahmed, Brendan Spooner, John Isherwood, Mark Lane, Emma Orrock, and Ashley Dennison. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus*, 15(10), 2023.

Adi Akavia, Max Leibovich, Yehezkel S Resheff, Roey Ron, Moni Shahar, and Margarita Vald. Privacy-preserving decision trees training and prediction. *ACM Transactions on Privacy and Security*, 25(3): 1–30, 2022.

Martin Albrecht, Melissa Chase, Hao Chen, Jintai Ding, Shafi Goldwasser, Sergey Gorbunov, Shai Halevi, Jeffrey Hoffstein, Kim Laine, Kristin Lauter, et al. Homomorphic encryption standard. *Protecting privacy through homomorphic encryption*, pages 31–62, 2021.

Mikhail Babenko and Elena Golimblevskaia. Euclidean division method for the homomorphic scheme CKKS. In *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pages 217–220. IEEE, 2021.

Charles George Broyden. The convergence of a class of double-rank minimization algorithms. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

Justine B Bukabau, Ernest K Sumaili, Etienne Cavalier, Hans Pottel, Bejos Kifakiou, Aliocha Nkodila, Jean Robert R Makulo, Vieux M Mokoli, Chantal V Zinga, Augustin L Longo, et al. Performance of glomerular filtration rate estimation equations in congolese healthy adults: the inopportunity of the ethnic correction. *PloS one*, 13(3):e0193384, 2018.

Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Homomorphic encryption for arithmetic of approximate numbers. In *Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I 23*, pages 409–437. Springer, 2017.

Jung Hee Cheon, Dongwoo Kim, and Duhyeong Kim. Efficient homomorphic comparison methods with optimal complexity. In *Advances in Cryptology–ASIACRYPT 2020: 26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7–11, 2020, Proceedings, Part II 26*, pages 221–256. Springer, 2020.

Ping-yeh Chiang, Renkun Ni, David Yu Miller, Arpit Bansal, Jonas Geiping, Micah Goldblum, and Tom Goldstein. Loss landscapes are all you need: Neural network generalization can be explained without the implicit bias of gradient descent. In *The Eleventh International Conference on Learning Representations*, 2023.

Niels Johan Christensen, Samuel Demharter, Meera Machado, Lykke Pedersen, Marco Salvatore, Valdemar Stentoft-Hansen, and Miquel Triana Iglesias. Identifying interactions in omics data for clinical biomarker discovery using symbolic regression. *Bioinformatics*, 38(15):3749–3758, 2022.

Frank Cole. Pediatric formulas for the anesthesiologist. *AMA journal of diseases of children*, 94(6): 672–673, 1957.

Kelong Cong, Debajyoti Das, Jeongeun Park, and Hilder VL Pereira. Sortinghat: Efficient private decision tree evaluation via homomorphic encryption and transciphering. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 563–577, 2022.

Naveen Eipe, Nicholas Barrowman, Hilary Writer, and Dermot Doherty. A weight-based formula for tracheal tube size in children. *Pediatric Anesthesia*, 19(4):343–348, 2009.

Andres Elovic and Ali Pourmand. MDCalc medical calculator app review. *Journal of digital imaging*, 32:682–684, 2019.

Junfeng Fan and Frederik Vercauteren. Somewhat practical fully homomorphic encryption. *Cryptology ePrint Archive*, 2012.

Candida Ferreira. Gene expression programming in problem solving. *Soft Computing and Industry: Recent Applications*, pages 635–653, 2002.

Kei Sen Fong and Mehul Motani. DistilSR: A distilled version of gene expression programming symbolic regression. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, pages 567–570, 2023.

Kei Sen Fong and Mehul Motani. Multi-level symbolic regression: Function structure learning for multi-level data. In *International Conference on Artificial Intelligence and Statistics*, pages 2890–2898. PMLR, 2024.

Kei Sen Fong and Mehul Motani. Symbolic regression enhanced decision trees for classification tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12033–12042, 2024.

Kei Sen Fong, Shelvia Wongso, and Mehul Motani. Rethinking symbolic regression: Morphology and adaptability in the context of evolutionary algorithms. In *The Eleventh International Conference on Learning Representations*, 2023.

Robert Hooke and Terry A Jeeves. 'Direct Search' solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961.

Lesley A Inker, Christopher H Schmid, Hocine Tighiouart, et al. Estimating glomerular filtration rate from serum creatinine and cystatin c. *New England Journal of Medicine*, 367(1):20–29, 2012.

Lesley A Inker, Nwamaka D Eneanya, Coresh, et al. New creatinine-and cystatin c–based equations to estimate gfr without race. *New England Journal of Medicine*, 385(19):1737–1749, 2021.

Pierre-Alexandre Kamienny, Guillaume Lample, Sylvain Lamprier, and Marco Virgolin. Deep generative symbolic regression with monte-carlo-tree-search. *Proceedings of the 40th International Conference on Machine Learning*, 2023.

Bangul Khan, Hajira Fatima, Ayatullah Qureshi, Sanjay Kumar, Abdul Hanan, Jawad Hussain, and Saad Abdullah. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomedical Materials & Devices*, pages 1–8, 2023.

Andrey Kim, Antonis Papadimitriou, and Yuriy Polyakov. Approximate homomorphic encryption with reduced approximation error. In *Cryptographers' Track at the RSA Conference*, pages 120–144. Springer, 2022.

Hyeonsik Kim, Hyun-Kyu Yoon, Hyeonhoon Lee, Chul-Woo Jung, and Hyung-Chul Lee. Predicting optimal endotracheal tube size and depth in pediatric patients using demographic data and machine learning techniques. *Korean Journal of Anesthesiology*, 76(6):540, 2023.

Miran Kim, Yongsoo Song, Shuang Wang, Yuhou Xia, Xiaoqian Jiang, et al. Secure logistic regression based on homomorphic encryption: Design and evaluation. *JMIR medical informatics*, 6(2): e8805, 2018.

John R Koza. Genetic programming. on the programming of computers by means of natural selection. *Complex adaptive systems*, 1992.

William La Cava, Patryk Orzechowski, Bogdan Burlacu, Fabrício Olivetti de França, Marco Virgolin, Ying Jin, Michael Kommenda, and Jason H Moore. Contemporary symbolic regression methods and their relative performance. *Neurips Track on Datasets and Benchmarks.*, 2021.

I Lawrence and Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, pages 255–268, 1989.

Joon-Woo Lee, Hyungchul Kang, Yongwoo Lee, Woosuk Choi, Jieun Eom, Maxim Deryabin, Eunsang Lee, Junghyun Lee, Donghoon Yoo, Young-Sik Kim, et al. Privacy-preserving machine learning with fully homomorphic encryption for deep neural network. *iEEE Access*, 10:30039–30054, 2022.

Andrew S Levey and Lesley A Inker. GFR as the "gold standard": estimated, measured, and true. *American Journal of Kidney Diseases*, 67(1):9–12, 2016.

Andrew S Levey, Josef Coresh, Tom Greene, Lesley A Stevens, Yaping Zhang, Stephen Hendriksen, John W Kusek, Frederick Van Lente, and Chronic Kidney Disease Epidemiology Collaboration*. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Annals of internal medicine*, 145(4):247–254, 2006.

Andrew S Levey, Lesley A Stevens, Christopher H Schmid, Yaping Zhang, Alejandro F Castro III, Harold I Feldman, John W Kusek, Paul Eggers, Frederick Van Lente, Tom Greene, et al. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*, 150(9):604–612, 2009.

T Nathan Mundhenk, Mikel Landajuela, Ruben Glatt, Claudio P Santiago, Daniel M Faissol, and Brenden K Petersen. Symbolic regression via neural-guided genetic programming population seeding. *Advances in Neural Information Processing Systems*, 2021.

Kundan Munjal and Rekha Bhatia. A systematic review of homomorphic encryption and its contributions in healthcare industry. *Complex & Intelligent Systems*, 9(4):3759–3786, 2023.

Samanvaya Panda. Principal component analysis using CKKS homomorphic scheme. In *Cyber Security Cryptography and Machine Learning: 5th International Symposium, CSCML 2021, Be'er Sheva, Israel, July 8–9, 2021, Proceedings 5*, pages 52–70. Springer, 2021.

Brenden K Petersen, Mikel Landajuela Larma, T Nathan Mundhenk, Claudio P Santiago, Soo K Kim, and Joanne T Kim. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. *The International Conference on Learning Representations*, 2019.

Thomas Prantl, Lukas Horn, Simon Engel, Lukas Iffländer, Lukas Beierlieb, Christian Krupitzer, André Bauer, Mansi Sakarvadia, Ian Foster, and Samuel Kounev. De bello homomorphico: Investigation of the extensibility of the openfhe library with basic mathematical functions by means of common approaches using the example of the ckks cryptosystem. *International Journal of Information Security*, pages 1–21, 2023.

Marc Raynaud, Solaf Al-Awadhi, Ivana Juric, Gillian Divard, Yannis Lombardi, Nikolina Basic-Jukic, Olivier Aubert, Laurence Dubourg, Ingrid Masson, Christophe Mariat, et al. Race-free estimated glomerular filtration rate equation in kidney transplant recipients: development and validation study. *bmj*, 381, 2023.

Kanthi Sarpatwar, Nalini K Ratha, Karthik Nandakumar, Karthikeyan Shanmugam, James T Rayfield, Sharath Pankanti, and Roman Vaculin. Privacy enhanced decision tree inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–35, 2020.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324 (5923):81–85, 2009.

George J Schwartz, Alvaro Mun, Michael F Schneider, et al. New equations to estimate GFR in children with CKD. *Journal of the American Society of Nephrology*, 20(3):629–637, 2009.

Ming-Hung Shih, Chin-Yang Chung, Bai-Chuan Su, Chao-Tsen Hung, Shu-Yam Wong, TK Wong, et al. Accuracy of a new body length-based formula for predicting tracheal tube size in chinese children. *Chang Gung Med J*, 31(3):276–279, 2008.

Anas Shikha and Asem Kasem. The development and validation of artificial intelligence pediatric appendicitis decision-tree for children 0 to 12 years old. *European Journal of Pediatric Surgery*, 33(05): 395–402, 2023.

Neeloofar Soleimanpour and Maralyssa Bann. Clinical risk calculators informing the decision to admit: A methodologic evaluation and assessment of applicability. *Plos one*, 17(12):e0279294, 2022.

Satheeskumar Subramani, Maitree Pandey, Anshu Gupta, Pramod Kohli, and Preeti Goyal Varshney. Comparative study of different formulae for prediction of best fit endotracheal tube size in children. *Ain-Shams Journal of Anesthesiology*, 15(1): 1–8, 2023.

Alexis A Topjian, Tia T Raymond, Dianne Atkins, Melissa Chan, Jonathan P Duff, Benny L Joyner Jr, Javier J Lasa, Eric J Lavonas, Arielle Levy, Melissa Mahgoub, et al. Part 4: pediatric basic and advanced life support: 2020 american heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*, 142(16_Suppl_2):S469–S523, 2020.

Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

Haishuai Wang, Benjamin Bowe, Zhicheng Cui, Hong Yang, S Joshua Swamidass, Yan Xie, and Ziyad Al-Aly. A deep learning approach for the estimation

of glomerular filtration rate. *IEEE Transactions on NanoBioscience*, 21(4):560–569, 2022.

Pawan Whig, Shama Kouser, Ashima Bhatnagar Bhatia, Rahul Reddy Nadikattu, and Pavika Sharma. Explainable machine learning in healthcare. In *Explainable Machine Learning for Multimedia Based Healthcare Applications*, pages 77–98. Springer, 2023.

Casper Wilstrup and Chris Cave. Combining symbolic regression with the cox proportional hazards model improves prediction of heart failure deaths. *BMC Medical Informatics and Decision Making*, 22 (1):1–7, 2022.

Jean-Baptiste Woillard, Charlotte Salmon Gandonnière, Alexandre Destere, Stephan Ehrmann, Hamid Merdji, Armelle Mathonnet, Pierre Marquet, and Chantal Barin-Le Guellec. A machine learning approach to estimate the glomerular filtration rate in intensive care unit patients based on plasma iohexol concentrations and covariates. *Clinical Pharmacokinetics*, 60:223–233, 2021.

Alexander Wood, Kayvan Najarian, and Delaram Kahrobaei. Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Computing Surveys (CSUR)*, 53(4):1–35, 2020.

Kexin Xu, Benjamin Hong Meng Tan, Li-Ping Wang, Khin Mi Mi Aung, and Huaxiong Wang. Privacy-preserving outsourcing decision tree evaluation from homomorphic encryption. *Journal of Information Security and Applications*, 77:103582, 2023.

Masahiro Yagisawa. Fully homomorphic encryption without bootstrapping. *Cryptology ePrint Archive*, 2015.

Zihan Zhou, Wenjie Guo, Dingqi Liu, Jose Ramon Nsue Micha, Yue Song, and Shuhua Han. Multiparameter prediction model of immune checkpoint inhibitors combined with chemotherapy for non-small cell lung cancer based on support vector machine learning. *Scientific Reports*, 13(1):4469, 2023.

Pei-Er Zhuang, Jiang-Hong Lu, Wei-Kai Wang, and Ming-Hua Cheng. A new formula based on height for determining endotracheal intubation depth in pediatrics: A prospective study. *Journal of Clinical Anesthesia*, 86:111079, 2023.

Appendix A. Additional Information for CKKS

Table 10: Scaling of Precision & Time Taken with respect to Equation complexity, k (Kim et al., 2022)

Equation form	Equation complexity, k	Precision (in bits)	Time Taken
$\prod_{i=1}^{2^k} \mathbf{x}_i$	1	21.8	4.01 ms
	2	20.1	34.25 ms
	3	20.7	0.1 s
	4	17.8	0.29 s
	5	17.3	0.73 s
	6	15.9	1.78 s
	7	14.3	8.86 s
$\sum_{i=0}^{k} \mathbf{x}^i$	2	21.8	4.14 ms
	4	19.4	29.39 ms
	8	19.1	75.14 ms
	16	16.9	0.17 s
	32	16.3	0.38 s
	48	15.1	0.67 s
	64	14.9	0.82 s

Table 11: Security level against CKKS parameters (i.e. Polynomial modulus degree and sum of bit sizes of the coefficient modulus) from the Homomorphic Encryption Standard (Albrecht et al., 2021)

Polynomial modulus degree	Sum of bit sizes of the coefficient modulus	Security level (in bits)
1024	27	128
	19	192
	14	256
2048	54	128
	37	192
	29	256
4096	109	128
	75	192
	58	256
8192	218	128
	152	192
	118	256
16384	438	128
	305	192
	237	256
32768	881	128
	611	192
	476	256

CKKS is the state-of-the-art homomorphic encryption scheme for machine learning (Lee et al., 2022; Panda, 2021), with the key unique strength being that it deals with encrypted real numbers naturally without requiring an explicit quantization step. Recent developments on encrypted machine learning have been done with CKKS, such as work done by Panda (2021) and Akavia et al. (2022). The unique strength of CKKS in supporting real number arithmetic, which is not readily available in other homomorphic encryption schemes. Beyond homomorphic encryption, methods like (i) multiparty computation requires the strong assumption of collaboration between multiple parties, which may not be practical or feasible and introduces complexity and overhead associated with coordinating multiple parties and (ii) differential privacy focuses on protecting against re-identification attacks but may not prevent all forms of privacy breaches. It is precisely the uniqueness of homomorphic encryption in allowing inherent arithmetic operations on encrypted raw data, that gives it the widely-known title of being the "holy grail of encryption" (Wood et al., 2020).

CKKS is built upon the hardness of the Ring Learning with Errors (RLWE) problem. RLWE is the task of recovering a secret polynomial modulo an irreducible polynomial when given noisy samples of the polynomial at random points. The difficulty of solving RLWE lies in distinguishing between random noise and structured information about the secret polynomial. The security of CKKS relies on the assumption that solving the RLWE problem is computationally difficult, even for adversaries equipped with significant computational resources. For the algorithmic steps and operations, we direct the reader to the original work by Cheon et al. (2017), Section 3.3., in particular on the 5 operations: KeyGen, Enc, Dec, Add, Mult.

Appendix B. Additional Information for K-Expressions

K-Expressions are fixed-length strings that can be decoded to form variable-length expression. The decoding process is done by reading the K-Expression starting from the left-most symbol of the string an building an expression tree by filling up empty spots (with top-most then left-most priority), until the tree is filled Ferreira (2002). For example, the string "$*+-abcde$" is decoded as $(a+b)*(c-d)$. In this case, the symbol e being in excess and not included in the already full expression. Thus, the length of the K-Expression need not be equal to the length of the expression tree, enables the creation of expressions of variable size from fixed length string representation. K-Expressions also have the To ensure that all *K-expressions* decode to form a valid mathematical expression, it is necessary that *K-expressions* have a tail component, in which only *terminal symbols* are present Ferreira (2002). By utilizing K-Expressions, DistilSR (Fong and Motani, 2023) achieved the best performance on datasets with compact ground truth equations, outperforming other SR methods (which bloats easily by evolving long equations).

Appendix C. Statistics on Existing Medical Equations

Based on the distribution of equations based on 53 medical specialities from MDCalc (consisting of 725 equations for over 200 diseases as seen in Figure 3), the top 15 specialities are: Internal Medicine (12.564%), Hospitalist Medicine (11.379%), Emergency Medicine (11.965%), Critical Care (11.001%), Family Practice (11.164%), Primary Care (9.605%), Gastroenterology (8.59%), Hematology and Oncology (8.8%), Cardiology (8.755%), Neurology (8.094%), General Surgery (8.665%), Pulmonology (8.709%), Pediatrics (7.155%), Geriatrics (6.239%), Hepatology (6.164%).

Additionally, based on Clarivate's Web of Science, as seen in Figure 4 the top 15 areas with publications on medical equations are: Oncology (11.249%), Medicine General Internal (11.085%), Cardiac Cardiovascular Systems (8.852%), Surgery (8.542%), Clinical Neurology (4.897%), Gastroenterology Hepatology (4.252%), Pharmacology Pharmacy (4.005%), Urology Nephrology (3.615%), Medicine Research Experimental (3.589%), Radiology Nuclear Medicine Medical Imaging (3.382%), Nutrition Dietetics (3.265%), Peripheral Vascular Disease (3.206%), Endocrinology Metabolism (3.066%), Pediatrics (2.782%), Orthopedics (2.678%).

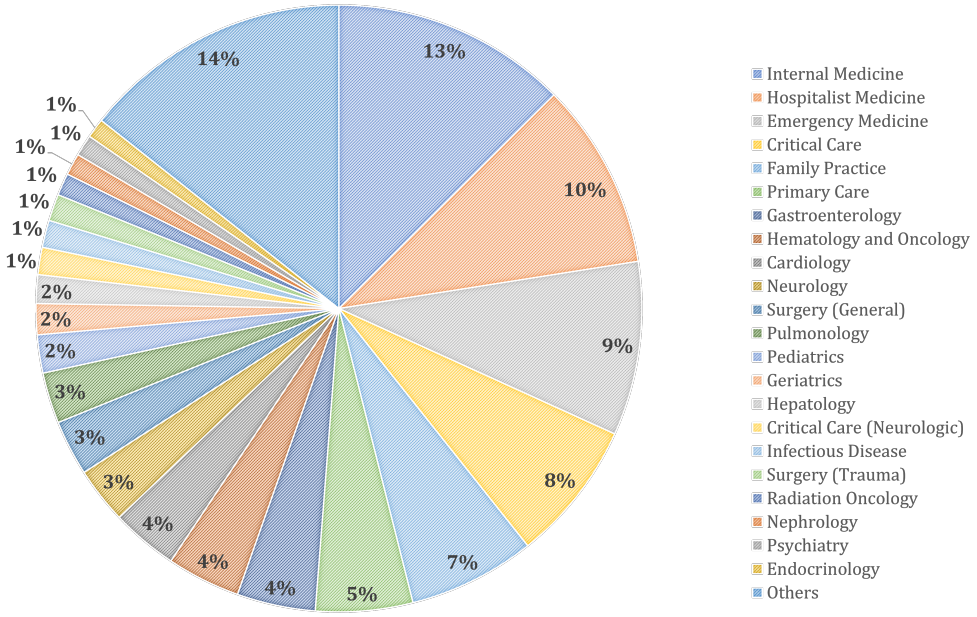


Figure 3: Distribution of medical equations across 53 medical specialities from MDCalc.

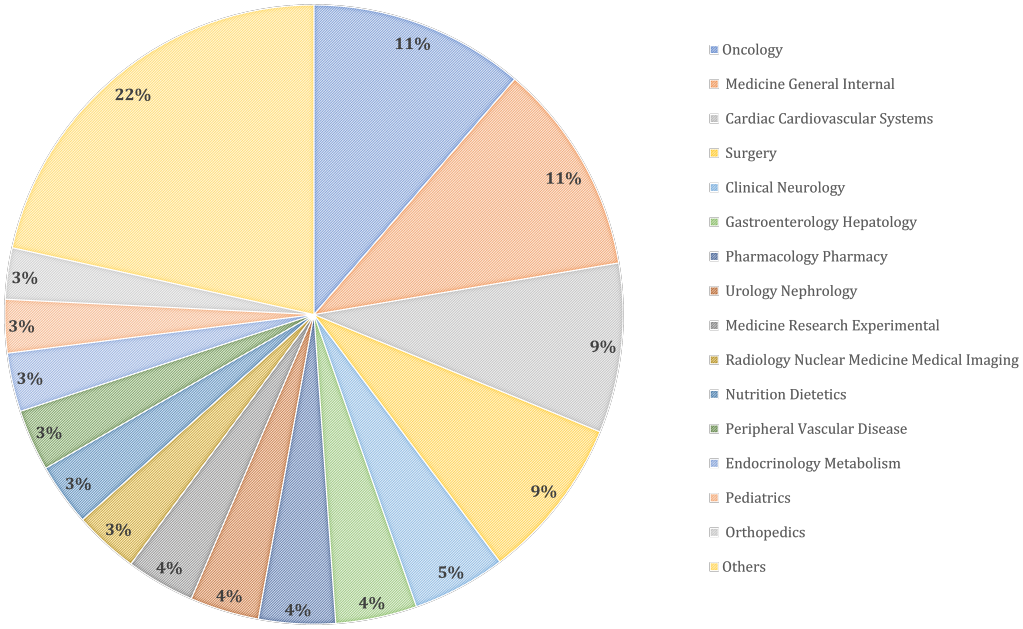
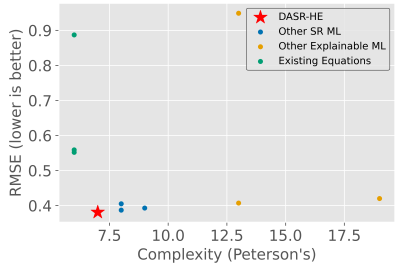
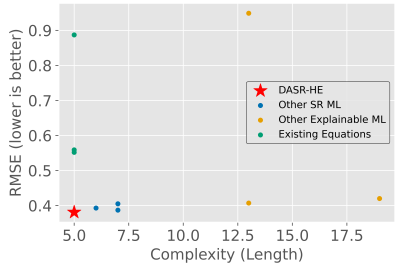


Figure 4: Distribution of publications on medical equations from Clarivate's Web of Science.

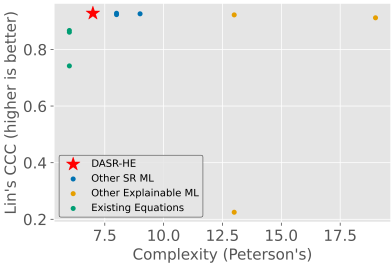
Appendix D. Other Performance-Complexity Graphs



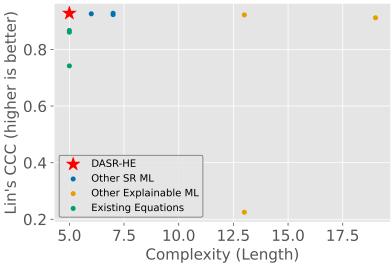
(a) RMSE vs Peterson's.



(b) RMSE vs Length.

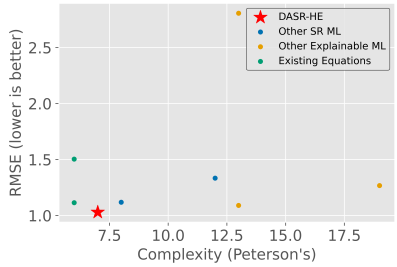


(c) CCC vs Peterson's.

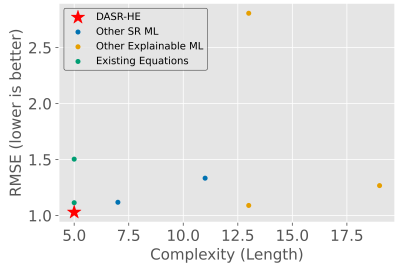


(d) CCC vs Length.

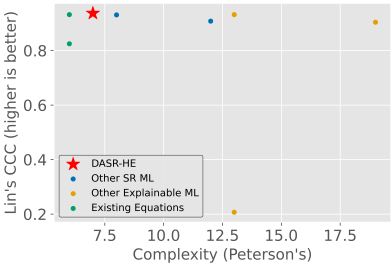
Figure 5: Plots of various prediction metrics against various complexity metrics for ETT Internal Diameter. DASR-HE pareto dominates all other approaches, in terms of performance and complexity.



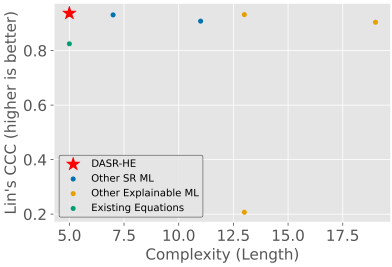
(a) RMSE vs Peterson's.



(b) RMSE vs Length.



(c) CCC vs Peterson's.



(d) CCC vs Length.

Figure 6: Plots of various prediction metrics against various complexity metrics for ETT Depth. DASR-HE pareto dominates all other approaches, in terms of performance and complexity.