European Journal of Mechanics / A Solids 114 (2025) 105780

Contents lists available at ScienceDirect

European Journal of Mechanics / A Solids

journal homepage: www.elsevier.com/locate/ejmsol

Full Length Article

A Transfer Learning method for deep drawing force prediction of sheet metal

Yingjian Guo [a], Gregor Kosec [b], Lihua Wang [c,*], Magd Abdel Wahab [a,b,d,**]

[a] Faculty of Engineering and Architecture, Ghent University, Technologiepark Zwijnaarde 46, Zwijnaarde, B-9052, Belgium
[b] Parallel and Distributed Systems Laboratory, Jožef Stefan Institute, Jamova cesta 39, SI-1000, Ljubljana, Slovenia
[c] School of Aerospace Engineering and Applied Mechanics, Tongji University, Shanghai, 200092, China
[d] College of Engineering, Yuan Ze University, Taiwan, Republic of China

A R T I C L E I N F O

Keywords:
Transfer Learning
Deep drawing force
Deep neural network model
Metal forming
Finite element method

A B S T R A C T

The deep drawing process involves complex deformations with highly nonlinear interactions among geometry, physics, and boundary conditions. Deep Drawing Force (DDF) refers to the force exerted by the punch on the sheet during forming and is essential for mechanical design and the prediction of failure modes such as wrinkling or fracture. This study proposes a data-driven transfer learning framework to predict DDF under various process conditions, where knowledge learned from analytically generated data is transferred to a model trained on limited numerical or experimental data, effectively reducing the data requirement while maintaining high prediction accuracy. The transfer learning in this work is entirely data-driven and relies on analytically generated data to pre-train a base model. Specifically, a large synthetic dataset is generated for pre-training using theoretical forming force equations based on the energy method. This base model is then fine-tuned using a small number of high-fidelity Finite Element Method (FEM) results that have been experimentally validated, finally, the prediction performance is evaluated by testing on unknown FEM data. The pre-trained knowledge is also transferred to tasks involving parts with different geometries, demonstrating strong generalization capability. Experimental results show that the transfer learning framework significantly improves prediction accuracy while reducing data requirements, offering an efficient and cost-effective solution for deep drawing process modelling.

Nomenclature

ANA	Analytical solution
a	The horizontal distance between the centre of the punch radius and the centre of the die radius
B	material strength coefficient
c	The clearance between punch and die
DDF	Deep drawing force
DNN	Deep neural network
D_p	Diameter of punch
EXP	Experiment
FEM	Finite element method
F_b	Blank holder force
F_p	Punch force
h	Punch distance
M	The bending moment
MAE	Mean Absolute Error
ML	Machine learning

(continued on next column)

(*continued*)

n	Strain hardening exponent
N_b	Bending work
N_e	The work by the external force
N_f	Friction work
N_p	Plastic deformation work
R_0, R	Initial and current radius of outer flange
R_C	The radius of the moving boundary points of region BC and region CD
R_D	Radius to the die lip
r	Thickness anisotropy coefficient
r_{dc}	Die corner radius
S_f	The friction surface
t	Time
u	The radial displacement
$\dot{u}$	The velocity of the point
V	The deformation body volume
ν	The drawing rate
α	Contact angle

(continued on next page)

* Corresponding author.
** Corresponding author. Faculty of Engineering and Architecture, Ghent University, Technologiepark Zwijnaarde 46, Zwijnaarde, B-9052, Belgium.
E-mail addresses: lhwang@tongji.edu.cn (L. Wang), agd.abdelwahab@UGent.be (M. Abdel Wahab).

https://doi.org/10.1016/j.euromechsol.2025.105780
Received 9 April 2025; Received in revised form 28 June 2025; Accepted 5 July 2025
Available online 7 July 2025

(*continued*)

$\varepsilon_r, \varepsilon_\theta, \varepsilon_t$	Plastic strain in radial, circumferential, and thickness directions
$\overline{\varepsilon}$	Equivalent plastic strain
$\dot{\overline{\varepsilon}}$	The equivalent strain rate
$\dot{\varepsilon}_r, \dot{\varepsilon}_\theta, \dot{\varepsilon}_t$	Plastic strain rate in radial, circumferential, and thickness directions
θ	The angle between the radial velocity and the actual velocity directions
μ	Coefficient of friction
ρ, x	Horizontal coordinate of the point before and after deformation
$\overline{\sigma}$	The equivalent stress
σ_N	Normal stress
σ_s	Yield stress
τ	The contact surface friction force
$\dot{\varphi}$	Bending angular velocity

1. Introduction

Deep drawing is a metal sheet stamping process in which a blank holder secures the sheet metal while a punch draws it into a die to form the desired shape. It is commonly used to manufacture hollow components with complex curves.

The deep drawing process reduces the need for expensive machining operations and improves production efficiency. It can be combined with other forming techniques to fabricate metal parts with complex shapes, lightweight structures, and excellent mechanical properties. As a result, deep drawing is extensively utilized in critical industrial fields, including aerospace, automotive manufacturing, home appliances, micro-components, and high-volume consumer goods production (Vollertsen et al., 2004). However, the deep drawing process is influenced by numerous process parameters, including material properties, such as elastic modulus, yield stress, and anisotropy (Zhao et al., 2024), geometric parameters, such as punch diameter (Guo et al., 2025), and die corner radius (Giorleo and Ceretti, 2022) and process conditions, such as blank holder force (Yoshihara et al., 2005), temperature (Zheng et al., 2017), and friction coefficient (Kim et al., 2007). These parameters significantly influence the forming quality in deep drawing. Excessive plastic deformation may cause localized thinning, leading to necking and fracture (Shutov et al., 2015; Basak et al., 2020; Basak and Panda, 2023; Firouzjaei et al., 2024). Another common failure mode is wrinkling, which occurs due to plastic buckling caused by high circumferential compressive stress (Du et al., 2021).

Deep Drawing Force (DDF) is the force applied by the punch to the sheet during the deep drawing process, progressively deforming the metal sheet and drawing it into the die to achieve the desired shape. DDF is a critical process parameter in deep drawing, used for calculating deformation work, determining mechanical equipment requirements, and designing tooling. DDF is utilized to determine the critical DDF for fracture instability (Leu, 1997) and the limiting drawing ratio (Leu, 1999; Fazli and Arezoo, 2012), thereby defining the maximum deep drawing height (Wan et al., 2001). Additionally, DDF is employed to assess the instability mechanisms leading to wrinkling in deep-drawn metal sheets. In industrial applications, blank holder force is an easily controllable parameter that influences DDF. Establishing the relationship between blank holder force and DDF allows for optimization to prevent failure (Zhao et al., 2004). Moreover, DDF is also used for real-time identification of material properties (Zhao and Wang, 2005). Therefore, accurately and effectively determining DDF is crucial for guiding the forming of sheet metal components. However, the deep drawing process involves a multitude of parameters, and during forming, the sheet is subjected to various forces, including bending, tension, friction, and compression, which collectively dictate the final product quality. Consequently, accurately and efficiently analysing DDF is both an essential and demanding task.

In the past, researchers have extensively used experimental, analytical, and numerical methods to study deep drawing problems. Experimental methods are costly and time-consuming, and their results are typically used to validate the accuracy of analytical and numerical approaches (Irthiea et al., 2014). Analytical methods are cost-effective and easy to implement, deriving analytical expressions based on theoretical analysis. Fereshteh-Saniee and Montazeran, (2003) applied three analytical equations to estimate the maximum DDF, determining that the Siebel formula yielded the most precise predictions. Choi et al. (2007) developed an analytical model and a FE model for hydraulic deep drawing conditions. The analytical model required approximately 30 s per calculation, whereas the FE simulation took 45 min. This comparison highlights the substantial computational efficiency advantage of analytical methods over time-consuming FEM simulations and repetitive experiments. Assempour and Taghipour, (2011) utilized analytical modelling grounded in incremental strain theory, leveraging mechanical and geometric relationships with the finite difference method to derive analytical equations for determining the punch force in hydraulic deep drawing. Rivas-Menchi et al. (2018) evaluated the accuracy of several analytical equations and found that their predictive performance varied depending on the number of parameters in the expressions, with prediction accuracy differing significantly under varying experimental conditions. Some of these analytical methods require intricate iterative computations, demanding continuous variable adjustments to meet error constraints, which can pose convergence challenges and require optimized algorithms to minimize computation time. To simplify calculations, certain assumptions and simplifications are often made, such as assuming constant blank thickness, simplifying friction treatment, and neglecting elastic deformation, which may reduce computational accuracy.

Sheet metal forming involves complex physical mechanisms, including nonlinearities caused by contact and friction, geometric nonlinearities, and nonlinear material behaviour. FEM is commonly used to address these challenges (Müller, M. et al., 2024; Dong et al., 2023; Luyen et al., 2022). Liu et al. (2019) conducted failure analysis on 6A16 aluminium alloy and investigated its deep drawing performance under varying blank holder forces using FEM. Choi and Huh (1999) analysed the deep drawing process of planar anisotropic materials using an improved membrane FE formulation. A comparison of deep drawing loads revealed that the modified membrane FE model for planar anisotropy provided more accurate results than the conventional membrane FE model for normal anisotropy. Kardan et al. (2018) optimized process parameters for deep drawing by combining FEM and experiments to minimize punch force and achieve a uniform thickness distribution. Their findings indicated that adjusting the initial blank thickness and die corner radius effectively reduced punch force and ensured a uniform thickness distribution. Atrian and Fereshteh-Saniee (2013) developed a FE model for deep drawing of laminated steel and brass sheets. Their study found that FE simulations overestimated DDF by approximately 10 % compared to experimental results. They also examined the effects of initial blank radius, friction conditions, and blank holder force on the maximum DDF. Sabet et al. (2021) investigated friction conditions of two aluminium alloys and incorporated surface roughness into the FE model. Their study revealed that an increase in blank holder force and friction coefficient led to a higher DDF. However, despite the high accuracy of FEM in addressing deep drawing problems involving multiple nonlinear behaviours, its application in deep drawing process design demands substantial computational resources (Cwiekala et al., 2011).

Globally, the adoption of intelligent manufacturing solutions driven by Machine Learning (ML) is accelerating to enhance productivity and reduce costs through predictive analytics. ML-based data-driven models have emerged as a viable alternative for addressing nonlinear challenges (Momeni et al., 2023). These models can capture the complex interactions of various process factors and have been applied across multiple fields, including material modelling (Bonatti and Mohr, 2021a), material design (Yang et al., 2021), and additive manufacturing (Cheng and De Waele, 2024). The rise of deep learning has opened new possibilities for intelligent manufacturing in sheet forming (El-Brawany et al., 2023; Samad et al., 2025; Link et al., 2025). Lee et al. (2024) propose a Gaussian process regression approach for deep-drawing blank design that captures nonlinear blank–form relationships and streamlines die development, and further reducing the cost of FEM simulations. Modanloo et al. (2025) propose a Gradient Enhanced–Expert Informed Neural Network (GE-EINN) model, which achieves accurate forming depth prediction using limited stamping data by integrating expert knowledge gradients, underscoring the value of hybrid ML strategies in data-scarce metal forming applications. Bonatti and Mohr (2021b) developed a FE-based Marciniak-Kuczynski model to generate a large database and constructed a neural network model for predicting the forming limit curves of sheet metal. Babu et al. (2010) developed an artificial neural network model to predict the maximum DDF, depth, and profile of tailor-welded blanks, achieving prediction errors within an acceptable range. Storm et al. (2024) developed a Graph Neural Network (GNN)-based approach to accelerate multiscale simulations of material mechanical responses, maintaining the multiscale nature of the problem while outperforming FE analysis in speed. Data-driven models offer high predictive efficiency; however, their performance is influenced by the size and quality of the dataset.

The Transfer Learning (TL) method offers a solution to data scarcity by transferring general knowledge learned from one dataset to a model trained on another dataset, thereby effectively enhancing generalization performance across different tasks (Yosinski et al., 2014). Duan et al. (2024) used GNN to predict incremental forming force, demonstrating the generality and effectiveness of TL methods in a limited dataset. Qu et al. (2023) generated a large amount of data using a phenomenological constitutive model to train a base model and then used TL to apply the pre-trained model to new tasks generated by numerical simulation with small datasets, showing that TL can effectively utilize limited data to improve prediction performance. Liu et al. (2020) used TL to transfer knowledge of easily obtainable electronic properties in semiconductors to help predict phonon properties, significantly reducing errors compared to direct neural network prediction. However, as far as the authors are aware, there has been no comprehensive attempt to apply TL to the deep drawing process.

The primary objective of this study is to develop a TL-based data-driven model for accurately and efficiently predicting DDF. This research aims to improve the accuracy of DDF prediction, reduce the reliance on expensive experiments, decrease the computational time needed for numerical simulations in design process, and utilize prior knowledge to minimize data requirements. Specifically, this study first employs an energy-based axisymmetric forming force theoretical formula to generate a dataset. This analytical approach considers multiple process parameters and incorporates certain assumptions to simplify calculations. While these assumptions introduce computational errors, they are useful for preliminary model training. Next, a high-fidelity dataset is created using an experimentally validated numerical model. TL is then utilized to incorporate the new data into the pre-trained model, enhancing its accuracy. Finally, knowledge transfer between different geometries is explored, specifically by reusing the trained model of cylindrical cups to predict DDF of rectangular components. Throughout the study, the test dataset is used to evaluate the model's predictive accuracy and computational efficiency.

2. Theoretical formulas and numerical method of deep drawing process

2.1. Theoretical model based on energy method

The drawing process of the cylindrical part is shown in Fig. 1. For the sake of calculation, it is assumed that the thickness remains constant during the deformation process. When considering the geometric relationships, the material thickness is neglected, and the neutral layer of the blank is considered as the research object of deformation, the single-side gap of the mould is distributed to the corner radius of the punch and die, thus, $r_{pc} = c/2 + r_p$, $r_{dc} = c/2 + r_d$ and $a = r_{pc} + r_{dc}$.

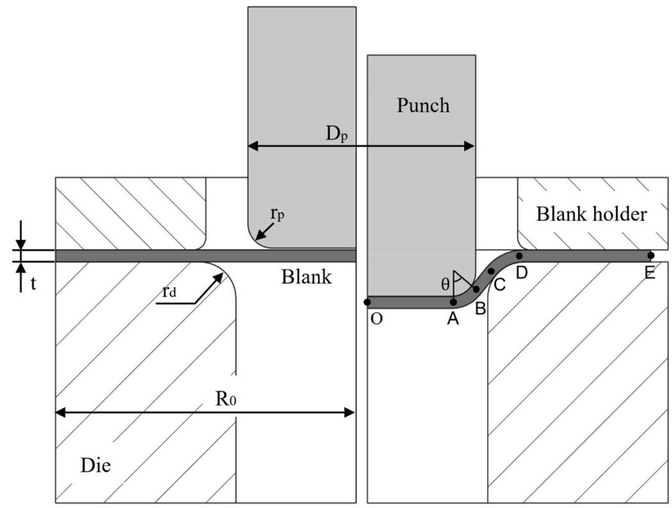

Fig. 1. Geometry of deep drawing assembly.

The part can be divided into five regions and simplified accordingly: 1) the region OA, which is in contact with the punch, 2) the punch fillet region AB, which bends at point B, 3) the free sidewall region BC, 4) the die fillet region CD and 5) the flange region DE. The last three regions undergo plastic deformation, with friction in regions CD and DE, and bending at points B, C and D.

For any deformation stage, the conservation of energy can be expressed as the work (N_e) done by the external force per unit time, which is equal to the sum of the plastic deformation work (N_p), bending work (N_b) and friction work (N_f) per unit time (Wang et al., 2017), that is:

$$\begin{cases} N_e = N_p + N_b + N_f \\ N_e = F_p \nu, N_p = \int_V \bar{\sigma}\dot{\bar{\varepsilon}} dV, N_b = M\dot{\varphi}, N_f = \int_{S_f} \tau \dot{u}_f dS_f \end{cases} \tag{2.1}$$

where F_p is the drawing force in kN, ν is the drawing rate in mm/s, and $\bar{\sigma}$, $\dot{\bar{\varepsilon}}$, V are the equivalent stress in the deformation body in MPa, equivalent strain rate in s^{-1} and deformation body volume in mm^3, respectively. M is the bending moment on the bending surface in N· mm and $\dot{\varphi}$ is the angular velocity of bending deformation in rad/s. τ is the contact surface friction force in N and $\dot{u}$ is the velocity of the point on the friction surface S_f in the opposite direction in mm/s.

Taking any point M on the plate before deformation, the horizontal coordinate of the point before and after deformation are ρ and x, respectively. When the punch distance is h, the radial displacement of the point is u, and the relationship $x = \rho + u(\rho, t)$ is obtained, where t is time, thus:

$$\dot{u} = \frac{\partial x}{\partial h}\frac{dh}{dt} = \frac{\partial x}{\partial h}\nu \tag{2.2}$$

The blank is subjected to friction in the die radius region (CD) and flange region (DE). Therefore, the velocity of the point in the region in the opposite direction of the friction force is:

$$\dot{u}_f = \dot{u}/\cos\theta \tag{2.3}$$

where θ is the angle between the radial velocity and the actual velocity directions.

The assumption of the constant volume condition and constant thickness leads to (Choi et al., 2007):

$$\varepsilon_t = -\varepsilon_r - \varepsilon_\theta = 0 \tag{2.4}$$

$$\varepsilon_r = -\varepsilon_\theta = \ln\left(\frac{\rho}{x}\right) \tag{2.5}$$

The expressions for equivalent strain $\overline{\varepsilon}$ and equivalent strain rate $\dot{\overline{\varepsilon}}$ are (Kim et al., 2008):

$$\overline{\varepsilon} = \frac{1+r}{\sqrt{1+2r}}\sqrt{\varepsilon_r^2+\varepsilon_\theta^2+\frac{2r}{1+r}\varepsilon_r\varepsilon_\theta} = \frac{\sqrt{2+2r}}{\sqrt{1+2r}}\ln\left(\frac{\rho}{x}\right) = 2\omega\ln\left(\frac{\rho}{x}\right) \tag{2.6}$$

$$\dot{\overline{\varepsilon}} = \frac{1+r}{\sqrt{1+2r}}\sqrt{\dot{\varepsilon}_r^2+\dot{\varepsilon}_\theta^2+\frac{2r}{1+r}\dot{\varepsilon}_r\dot{\varepsilon}_\theta} = \frac{\sqrt{2+2r}}{\sqrt{1+2r}}\frac{1}{x}\left|\frac{\partial x}{\partial h}\right|\nu = 2\omega\frac{1}{x}\left|\frac{\partial x}{\partial h}\right|\nu \tag{2.7}$$

$$\omega = \frac{\sqrt{1+r}}{\sqrt{2(1+2r)}} \tag{2.8}$$

where r is the thickness anisotropy coefficient of the plate.

Using the Holloman strain hardening equation, the equivalent stress can be written as:

$$\overline{\sigma} = \sigma_s + B\overline{\varepsilon}^n \tag{2.9}$$

where σ_s is the yield stress, B is the material strength coefficient and n is the strain hardening exponent.

According to the plastic bending theory of straight beams, the bending moment of the circumference should be $\frac{t^2}{4}\overline{\sigma}_i$, and the bending moment on the entire bending surface is:

region, i.e.

$$\begin{cases} \rho^2 = R_A^2 + 2r_{pc}\left[R_A\alpha - r_{pc}(\cos\alpha - 1)\right] + \dfrac{x^2-R_B^2}{\cos\alpha} & R_B \le x \le R_C \\ x = R_D - r_{dc}\sin\theta & \\ R_0^2 - \rho^2 = R^2 - R_D^2 + 2r_{dc}\left[R_D\theta + r_{dc}(\cos\theta - 1)\right] & R_C \le x \le R_D \\ R_0^2 - \rho^2 = R^2 - x^2 & R_D \le x \le R \end{cases} \tag{2.15}$$

By differentiating both sides of Equation (2.15) with respect to α, the following equation can be obtained:

$$\begin{cases} \left|\dfrac{\partial x}{\partial h}\right|_3 = \left|\dfrac{\partial x}{\partial \alpha}\right|_3 \dfrac{d\alpha}{dh} = \dfrac{\sin\alpha\cos\alpha}{2x}\dfrac{\left(x^2-R_B^2\right)}{R_C-R_B} & R_B \le x \le R_C \\ \left|\dfrac{\partial x}{\partial h}\right|_4 = \left|\dfrac{\partial x}{\partial \theta}\cdot\dfrac{d\theta}{dR}\cdot\dfrac{dR}{dh}\right| = \dfrac{\sin\alpha\cos\theta}{2}(R_c+R_B)\dfrac{\cos\theta}{R_D-r_{dc}\sin\theta} & R_C \le x \le R_D \\ \left|\dfrac{\partial x}{\partial h}\right|_5 = \left|\dfrac{\partial x}{\partial R}\dfrac{dR}{dh}\right| = \dfrac{\sin\alpha}{2x}(R_c+R_B) & R_D \le x \le R \end{cases} \tag{2.16}$$

Substituting Equation (1) into Equation (2), the expression for the equivalent strain $\overline{\varepsilon}_i$ in each deformation zone can be obtained as:

$$\begin{cases} \overline{\varepsilon}_3 = \omega\ln\left\{\dfrac{1}{x^2}\left[R_A^2 + 2r_{pc}\left(R_A\alpha + r_{pc}(1-\cos\alpha)\right)\right] + \dfrac{1}{\cos\alpha}\left(1 - \dfrac{R_B^2}{x^2}\right)\right\} & R_B \le x \le R_C \\ \overline{\varepsilon}_4 = \omega\ln\left[\dfrac{1}{x^2}\left(R_0^2 - R^2 + R_D^2 - 2r_{dc}\left[R_D\theta - r_{dc}(1-\cos\theta)\right]\right)\right] & R_C \le x \le R_D \\ \overline{\varepsilon}_5 = \omega\ln\left[\dfrac{1}{x^2}\left(R_0^2 - R^2 + x^2\right)\right] & R_D \le x \le R \end{cases} \tag{2.17}$$

$$M_i = \frac{t^2}{4}\overline{\sigma}_i(2\pi R_i) = \frac{\pi}{2}t^2R_i\overline{\sigma}_i, (i = B, C, D) \tag{2.10}$$

Arranging the above formula and substituting it into formula (2.1), we get the formula for DDF as (Zhao et al., 1998):

$$F_p = 2\omega\int_V \frac{\overline{\sigma}}{x}\left|\frac{\partial x}{\partial h}\right|dV + \int_{S_f}\frac{\tau}{\cos\beta}\left|\frac{\partial x}{\partial h}\right|dS_f + \frac{\pi}{2}t^2\sum R_i\overline{\sigma}_i\frac{\dot{\varphi}_i}{\nu} \tag{2.11}$$

2.1.1. Plasticity equation

From the geometric relationship, the relationship between the depth h and the contact angle α is:

$$\alpha = \mathrm{tg}^{-1}\frac{h(2a-h)}{2a(a-h)} \tag{2.12}$$

Since the area remains unchanged before and after deformation, it can be concluded that (Wu, 2022):

$$R^2 = R_0^2 + R_D^2 - R_A^2 - \frac{R_C^2 - R_B^2}{\cos\alpha} - 2\left(r_{pc}R_A + r_{dc}R_D\right)\alpha + 2\left(r_{dc}^2 - r_{pc}^2\right)(1-\cos\alpha) \tag{2.13}$$

$$R_A = d/2 - r_{pc}, R_B = R_A + r_{pc}\sin\alpha, R_D = R_A + a, R_C = R_D - r_{dc}\sin\alpha \tag{2.14}$$

The relationship $x = x(\rho, h)$ between the instantaneous coordinates x, the original coordinates ρ, and the stroke h of the punch of points in each deformation region can be derived through the theory of constant

2.1.2. Friction model

Considering only the friction conditions in regions CD and DE, it is assumed that the friction force on the surface of DE is uniformly distributed with a value of $\tau_{DE} = \mu\sigma_N$. Region CD satisfies two boundary conditions $\tau_{CD}|_{x=R_C} = 0$, $\tau_{CD}|_{x=R_D} = \mu\sigma_N$, and its distribution is assumed as:

$$\tau_{CD} = \mu\sigma_N\frac{\sin(\alpha-\theta)}{\sin\alpha} \tag{2.18}$$

$$\sigma_N = \frac{F_p + F_b}{\pi {r_{pc}}^2}K(\alpha) \tag{2.19}$$

where $K(\alpha) = \dfrac{r_{dc}^2}{R^2 - {R_D}^2 + \alpha r_{dc}R_D - \frac{2}{3}r_{dc}^2(1-\cos\alpha) + \mu\left[r_{dc}R_D(1-\alpha\,\mathrm{ctg}\alpha) - \frac{2}{3}r_{dc}^2\frac{(1-\cos\alpha)^2}{\sin\alpha}\right]}$ (Gao et al., 2018)

2.1.3. Bending angular velocity

Bending angular velocity at each point during the deformation stage is given by:

$$\begin{cases} \dot{\varphi}_B = \dot{\varphi}_C = \dfrac{\mathrm{d}\alpha}{\mathrm{d}t} = \dfrac{\mathrm{d}\alpha}{\mathrm{d}h}\dfrac{\mathrm{d}h}{\mathrm{d}t} = \dfrac{\cos^2\alpha}{R_C - R_B}\nu \\ \dot{\varphi}_D = \dfrac{\dot{u}}{r_{dc}} = \dfrac{\sin\alpha}{2R_D r_{dc}}(R_c + R_B)\nu \end{cases} \tag{2.20}$$

2.1.4. Force-stroke relationship

By organizing the above equations and substituting them into Equation (2.1), the analytical expression for the relationship between drawing forming force and stroke is obtained as:

$$F_p = \frac{\pi t(R_B + R_C)\sin\alpha}{1 - \mu\dfrac{\sin\alpha}{r_{dc}^2}(R_B + R_C)\cdot\left(\dfrac{\sin\alpha}{1+\cos\alpha}r_{dc} + R - R_D\right)\cdot K(\alpha)} * H(\alpha) \tag{2.21}$$

Where:

$$H(\alpha) = 2\omega\left(\int_{R_B}^{R_C}\overline{\sigma}_3\frac{x^2 - R_B^2}{R_C^2 - R_B^2}\cdot\frac{dx}{x} + \int_0^{\alpha}\overline{\sigma}_4\frac{r_{dc}\cos\theta}{R_D - r_{dc}\sin\theta}d\theta + \int_{R_D}^{R}\overline{\sigma}_5\cdot\frac{dx}{x}\right) +$$
$$\mu F_b\frac{1}{2\pi tR}\left[1 + \frac{2R}{r_{dc}^2}\left(\frac{\sin\alpha}{1+\cos\alpha}r_{dc} + R - R_D\right)\cdot K(\alpha)\right] +$$
$$\frac{\mathrm{t}}{4r_{dc}}\left[\frac{2r_{dc}R_B}{R_C^2 - R_B^2}\cot\alpha\cdot\cos\alpha\cdot\left(\frac{R_B}{Rc}\overline{\sigma}_{3B} + \overline{\sigma}_{4C}\right) + \overline{\sigma}_{5D}\right] \tag{2.21a}$$

2.2. Numerical analysis

A FE model of the cylindrical part drawing process established in the commercial software ABAQUS is shown in Fig. 2, and an explicit dynamic time integration scheme is used to efficiently solve highly nonlinear contact and large-deformation problems. Due to the symmetry of the model, only a quarter model is established to reduce computation time. The sheet is meshed with S4R elements (4-node reduced integration with hourglass control) with a global mesh size of 1.3 mm and a total of 1683 meshes. They are widely used in metal forming simulations due to their computational efficiency and ability to capture in-plane deformation. The tools (punch, die, blank holder) are meshed with discrete rigid shell elements (R3D4, 3D, 4-node rigid elements), with its motion controlled by a reference point. Surface-to-surface contact was defined with friction modelled using a penalty formulation and a constant Coulomb coefficient. A structured mesh was used for the blank with refined mesh zones in punch corner expected to undergo severe plastic deformation.

To validate the FE model, experimental parameters from the literature (Choi and Huh, 1999) were adopted. The blank material used is aluminium-killed steel, characterized by a stress-strain relationship represented as $\overline{\sigma} = 561.92(\overline{\varepsilon} + 0.00941)^{0.266}$ (N/mm^2). The anisotropy coefficients (r) $r0$, $r45$, and $r90$ are 1.751, 1.521, and 2.082, respectively. The geometric parameters are illustrated in Fig. 1, and the process variables employed in the simulation are listed in Table 1.

Table 1
Process and geometry parameters of the cylindrical cup deep drawing process.

Parameters	Value
Radius of sheet blank ($R0$)	55 mm
Sheet thickness (t)	0.74 mm
Diameter of punch (D_p)	50 mm
The corner radius of die (r_d)	8 mm
The corner radius of punch (r_p)	8 mm
The clearance between punch and die (c)	1 mm
Friction coefficient (μ)	0.17
Blank holder force (F_b)	19613 N
Punch displacement (h)	30 mm

Rectangular deep drawing is a typical non-axisymmetric deep drawing process, compared to axisymmetric blanks, rectangular deep drawing introduces additional parameters such as corner and straight edge regions, complicating the deformation process. The concept of equivalent diameter has been proposed in the literature (Daxin et al., 2008; Medell-Castillo et al., 2013) to extend axisymmetric deep drawing theory to rectangular parts, proving a close relationship between deep drawing processes for different shapes. However, applying some theoretical formulas to calculate maximum forming force has not been ideal (Rivas-Menchi et al., 2018), likely due to the equivalent diameter fails to effectively capture the influence of additional parameters in rectangular deep drawing. Therefore, this paper uses ML to describe the complex nonlinear relationships. By using TL to combine the knowledge of cylindrical drawing with the rectangular drawing dataset generated by FEM, we can further explore the hidden connections between them. We adopted the process parameters reported in the literature (Choi and Huh, 1999) to construct the FE model, as shown in Fig. 3. The mesh unit types of the mould and plate are the same as cylindrical drawing, and the fillet areas of the mould are also refined and the blank is meshed with a global mesh size of 1.25 mm and a total of 3264 meshes. We generate 17 sets of forming force data with different friction coefficients and plate thicknesses. These data are used to further train the TL model to verify its applicability in complex non-axisymmetric deep drawing.

3. Machine learning models

3.1. Deep Neural Network (DNN)

A fully connected DNN is constructed to train and learn the relationship between process parameters and punch force. The model consists of an input layer, hidden layers, and an output layer. This pointwise approach allows DNN to learn the underlying mapping:

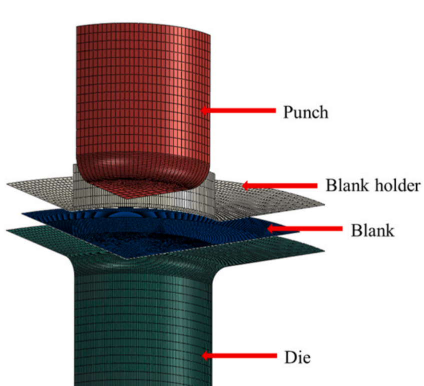

Fig. 2. FE model of cylindrical part deep drawing process.

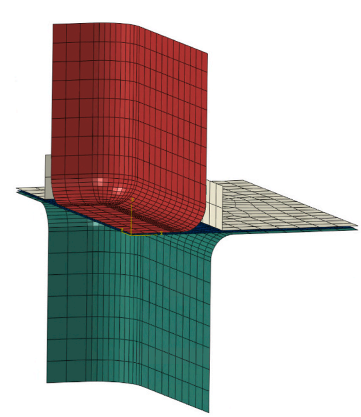

Fig. 3. FE model of rectangular plate deep drawing.

$$(r, t, h) \longmapsto F_p \tag{3.1}$$

where r is the coefficient of friction, t is the thickness, h is the stroke displacement and F_p is the force.

In an L-layer fully connected neural network, the l^{th} layer contains N_l neurons, and the output of the j^{th} neuron is the weighted average of the outputs from the previous $(l-1)$ layer, followed by the application of an activation function to introduce non-linearity. The complexity of the network architecture systematically increases from the input layer to the output layer, enabling it to effectively capture hierarchical features in the data. The general construction of an L-layer neural network is as follows (Haghighat et al., 2023):

$$A_j^l = f\left(\sum_{i=1}^{N_{l-1}} W_{ij}^l A_i^{l-1} + b_j^l\right) \tag{3.2}$$

where W_{ij}^l represents the matrix of weights and b_j^l denotes the biases associated with the l^{th} layer. $f(\cdot)$ is the activation function, enabling the network to model complex phenomena that are not linearly separable.

Common activation functions include RELU ($f(x) = \max(0, x)$), Sigmoid function ($f(x) = 1/(1+e^{-x})$), Tanh ($f(x) = (e^x - e^{-x})/(e^x + e^{-x})$) (Li et al., 2021). Also, DNN was trained using the Adam optimizer, a widely adopted algorithm for stochastic gradient-based optimization, Adam combines the benefits of adaptive learning rates from RMSProp and momentum-based gradient descent, making it suitable for efficiently training complex neural network models.

The performance of DNN is evaluated through the loss function, which estimates the difference between the predicted values and the target values. During training, the backpropagation algorithm is used to compute the gradient of the loss function with respect to the model weights, adjusting the weights in the direction that reduces the loss (Susmel, 2024). This iterative process continually improves the model's performance. Commonly used loss functions are: Mean Squared Error (Bonatti and Mohr, 2021b) ($MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$), Mean Absolute Error (Weng and Yuan, 2023) ($MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|$) and Smooth L1 Loss ($Smooth\ L1 = \begin{cases} |x| - 0.5, |x| > 1 \\ 0.5x^2, |x| < 1 \end{cases}$).

In ML and statistical modelling, evaluating the performance of a model is crucial. This paper uses the coefficient of determination (R^2) as a metric to assess the model's explanatory power over the data. The formula for calculating R^2 is as follows:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{3.3}$$

Where y_i is the target values, and $\widehat{y}_i$ is the predicted values, $\overline{y}$ is the average of all actual values, and n is the number of samples.

3.2. Transformer Learning (TL) method

3.2.1. TL theory

TL is a machine-learning paradigm in which knowledge gained from solving a source task is leveraged to improve learning performance on a related target task. Formally, the source domain and source task are as follows:

$$\mathscr{D}_s = \{\mathscr{X}_s, P_s(\mathbf{x})\}, \mathscr{T}_s = \{\mathscr{Y}_s, f_s(\mathbf{x})\} \tag{3.4}$$

where $\mathscr{X}_s$ is the feature space, $P_s(\mathrm{x})$ the marginal distribution, $\mathscr{Y}_s$ the label space, and f_s the predictive function. Similarly, the target domain and target task are defined as:

$$\mathscr{D}_t = \{\mathscr{X}_t, P_t(\mathbf{x})\}, \mathscr{T}_t = \{\mathscr{Y}_t, f_t(\mathbf{x})\} \tag{3.5}$$

In TL, $\mathscr{D}_s \neq \mathscr{D}_t$ or $\mathscr{T}_s \neq \mathscr{T}_t$ yet one seeks to learn f_t more effectively by transferring information from $(\mathscr{D}_s, \mathscr{T}_s)$.

3.2.2. Pre-training (source task)

In deep neural networks, TL most commonly takes the form of parameter transfer: a network is pre-trained on the source data to learn a hierarchy of features, where early layers capture general representations and later layers capture task-specific patterns (Pan and Yang, 2010). Based on the TL method, this paper uses an analytical equation based on the energy method to quickly generate a large amount of deep drawing data as a data set for pre-training the source model. Although the analytical solution has poor accuracy due to some empirical assumptions, it can still provide good trends and approximations. By updating all weights and biases, the basic DNN is trained on the source data set to minimize the source loss $\mathscr{L}_s$:

$$\{W^*, b^*\} = \arg\ min\ \mathscr{L}_s(W, b) \tag{3.6}$$

3.2.3. Fine-tuning (target task)

After obtaining the source model, the source model parameters are transferred to the target task, and the target data set for training the target model is formed using experimentally verified FE data or experimental data. A fine-tuning strategy is used to update the model parameters. During fine-tuning, one freezes the weights of the early layers ($1 \le l \le k$) and updates only the deeper layers ($k+1 \le l \le L$) on the target data to minimize the target loss $\mathscr{L}_t$. This strategy preserves general features while allowing specialized adaptation, reducing the amount of target data required to avoid overfitting (Yosinski et al., 2014).

Mathematical Formulation of Freezing & Fine-Tuning.

1. Frozen layers ($1 \le l \le k$):

$$W_{ij}^l = W_{ij}^{l^*}, b_j^l = b_j^{l^*} \tag{3.7}$$

2. Tuned layers ($k+1 \le l \le L$):

$$\left\{W_{ij}^l, b_j^l\right\} = \arg\ min\ \mathscr{L}_t\left(\left\{W_{ij}^{l^*}, b_j^{l^*}\right\}_{l \le k}, \left\{W_{ij}^l, b_j^l\right\}_{l > k}\right) \tag{3.8}$$

This paper adopts the TL method to predict deep drawing force. The source model is pre-trained based on the data calculated by the analytical method, the source model parameters are transferred to the target model, and the target model is fine-tuned using the experimentally verified FE data. The flow chart of TL is shown in Fig. 4. By reusing pre-trained features, TL can dramatically reduce data requirements, accelerate convergence, and improve generalization when target data are scarce or expensive. This makes it particularly well suited for engineering tasks, where obtaining large, high-fidelity labeled datasets, e.g., from experiments or detailed FEM simulations, is costly.

4. Validation of theoretical model and FEM

Based on the FE model established in Section 2.2, the variation of DDF with punch stroke is obtained, as shown in Fig. 5. The FEM and experimental results are in good agreement. Thus, the FE model established in this paper is completely reliable. In subsequent sections, this verified and reliable simulation model can be used to generate the data set required for ML.

A significant discrepancy exists between the results obtained from the analytical method and those from the experiment. The theoretical calculation formulas used in this study are based on the principle of energy conservation. While the calculation process takes various influencing factors into detailed consideration, certain assumptions and simplifications were made to simplify the analysis. These assumptions and simplifications include the following. First, to simplify the analysis, it is assumed that the average thickness of the blank remains constant

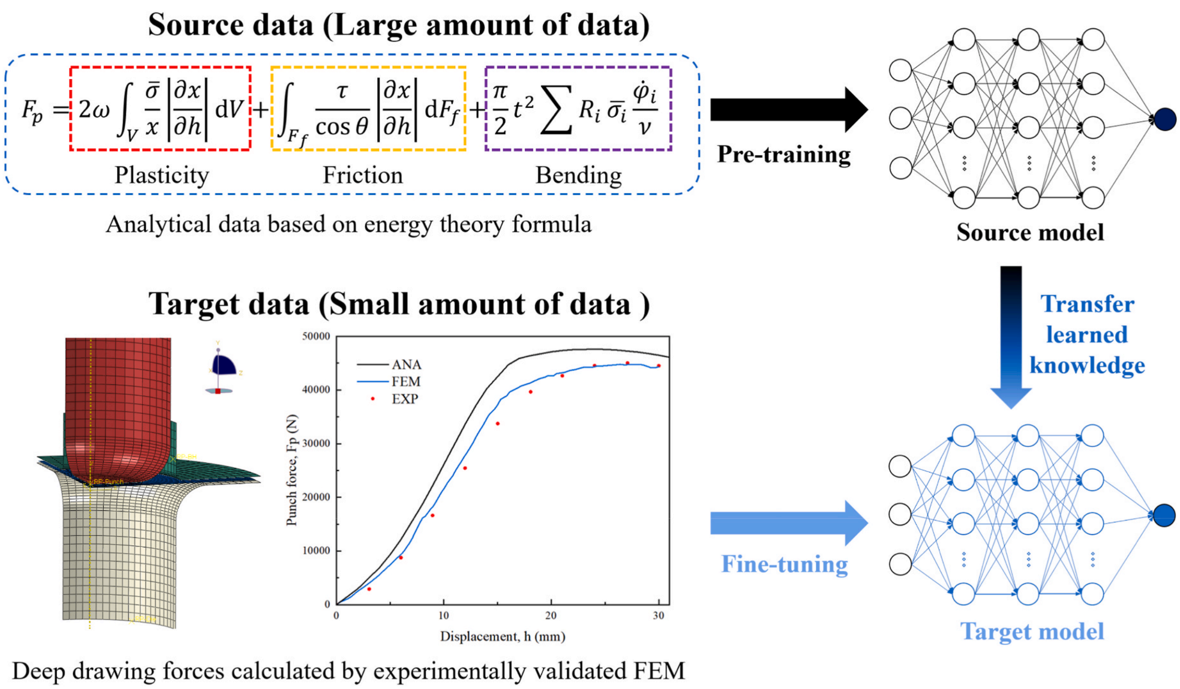

Fig. 4. TL schematic diagram of transferring knowledge from source model to target model.

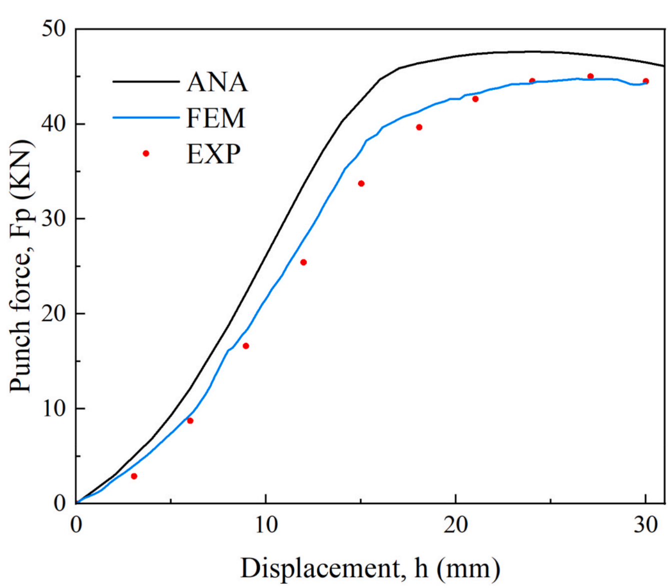

Fig. 5. Comparison between analytical solution, FEM and experiments (Choi and Huh, 1999).

before and after deformation, and the condition of constant plastic deformation volume is simplified to constant area. However, in actual deep drawing processes, the thickness of the blank does not change uniformly. Specifically, the blank at the punch bottom and corners thins, while at the die corners and outer edges, the thickness increases. This non-uniform thickness change is not fully accounted for in the theoretical model. Second, contact friction is also simplified. In actual deep drawing processes, the sheet's outer edge thickness gradually increases, leading to a continuously changing contact area. Additionally, as the outer edge thickens, the contact area in region DE decreases. However, to simplify the analysis, this study assumes a completely uniform distribution of contact normal stress in region DE, while region CD uses a continuity model based on boundary conditions to approximate the effect of friction. These simplifications may not fully capture the complex frictional behaviour in actual deep drawing processes. Additionally, this study neglects the elastic deformation of the metal material during the deep drawing process and its corresponding elastic energy. The elastic deformation of the metal affects the overall deformation behaviour and energy distribution. Ignoring this factor may lead to discrepancies between the theoretical model and actual conditions.

While these assumptions and simplifications facilitate theoretical analysis, they also introduce certain limitations. Nonetheless, these simplified theoretical calculations still capture the main features of the forming process. However, considering the above simplifications, future research may need to incorporate more complex numerical simulations or experimental data to further refine the model, providing a more accurate description of the complex physical phenomena in deep drawing processes.

4.1. Generation of database

The analytical solution presented in this article is an in-depth study of forming forces considering the influence of numerous parameters. A significant advantage is their ability to quickly calculate the drawing force for specified parameters. In this section, we use analytical solutions to generate extensive forming force-stroke data. Data-driven models are trained using artificially generated data and subsequently retrained using data from FE model. Several learning schemes based on TL are explored and compared.

The punch force generated by the verified and reliable FE model is used as high-fidelity data as a small batch target data set in TL. The effectiveness of TL directly depends on the similarity between the source data and the target data. The free parameters of the analytical equation source model are calibrated using the FE model.

Some simplifications have been made to the analytical equations, and it is understandable that the FE and the analytical solution data are slightly different. Despite some differences, the results show satisfactory agreement between the forming force-stroke data obtained from FE simulations and analytical equations. Subsequently, a large amount of forming force-stroke data was generated using the analytical equation for different friction coefficients and plate thicknesses for subsequent

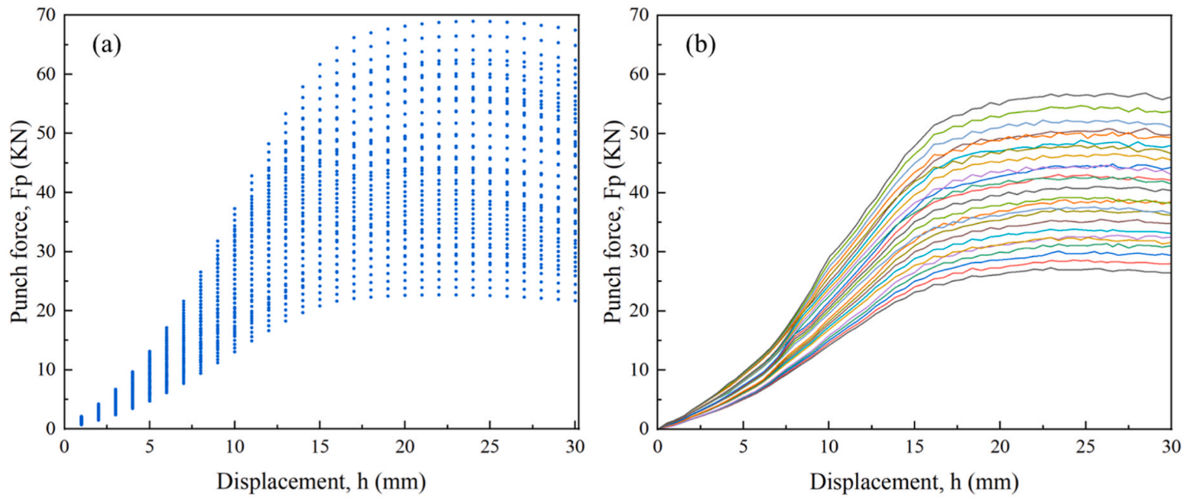

Fig. 6. DDFs generated by: (a) analytical solution and (b) FEM.

learning of pre-train model. We generated 49 different analytical force-displacement sequences covering 7 coefficients of friction (0.02–0.20, step size 0.03) and 7 sheet thicknesses (0.44–1.04 mm in step size 0.1 mm), as shown in Fig. 6(a). Each sequence is discretized into 30 data points, generating a rich source dataset for pre-training the source model. The target dataset was generated using the experimentally verified FE model, with a coefficient of friction in the range of 0.05–0.17 (interval 0.03) and a blank thickness range of 0.54 mm–0.94 mm (interval 0.1 mm), totally 25 groups, and each sequence is discretized into 33 data points, as shown in Fig. 6(b).

5. Results

5.1. Pre-trained model based on analytical equations

The hyperparameters of the pre-trained model were obtained through trial and error. DNN was configured with 3 hidden layers, each containing 40 neurons, a learning rate of 0.001, and a sigmoid activation function to introduce non-linearity. Prediction accuracy was evaluated using MAE, and the network was trained using the Adam optimizer. We select 5 sets of the 49 source datasets in Section 4.1 as test datasets, with the remaining data divided into training and validation sets in an 80 % and 20 % ratio, respectively. The training and validation losses are shown in Fig. 7(a) and 7(b) compares the test set results. R^2 is 0.9999, indicating that the pre-trained model effectively captured the features and patterns of the source dataset and maintained high predictive performance on unseen data.

5.2. Comparison with/without TL

After obtaining the pre-trained source model, we compare the performance of three modelling approaches, namely analytical, Non-TL model, and TL model, on the same FEM force–displacement dataset. Our TL workflow began with the pre-trained base network (Section 5.1) and fine-tuned all layers using only eight FEM-generated curves (33 discrete points each), reserving 17 unseen curves for testing. For a fair baseline, we also built a new DNN (Non-TL) from scratch on those eight curves, with hyperparameters optimized via Optuna (3 hidden layers, 87 neurons per layer, learning rate = 0.0073, batch size = 59).

Although the Non-TL achieves a respectable $R^2 = 0.957$ and MAE = 0.168, inspection of its prediction errors reveals systematic under- and over-estimation in the whole region (Fig. 8(b)), indicating that purely data-driven fitting struggles to capture subtle nonlinearities when data are scarce. By contrast, the TL model archive near-perfect metrics ($R^2 = 0.9995$, MAE = 0.016) and largely eliminates the bias patterns seen in the Non-TL results (Fig. 8(c)).

From the results in Fig. 8(a), it can be seen that the analytical equation significantly underestimates the larger deep drawing force, but

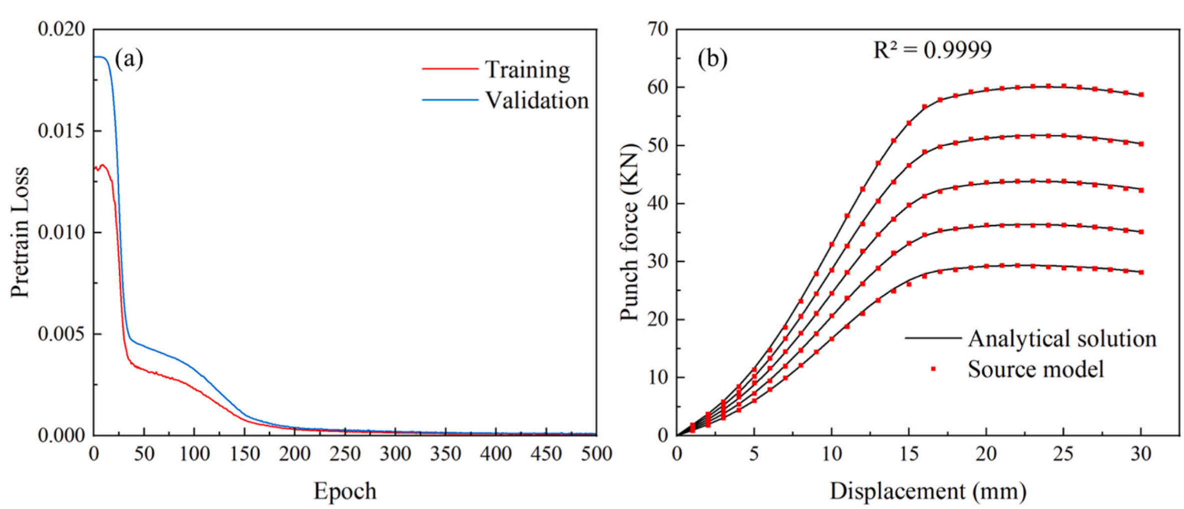

Fig. 7. (a) Loss function and (b) prediction results of the source model.

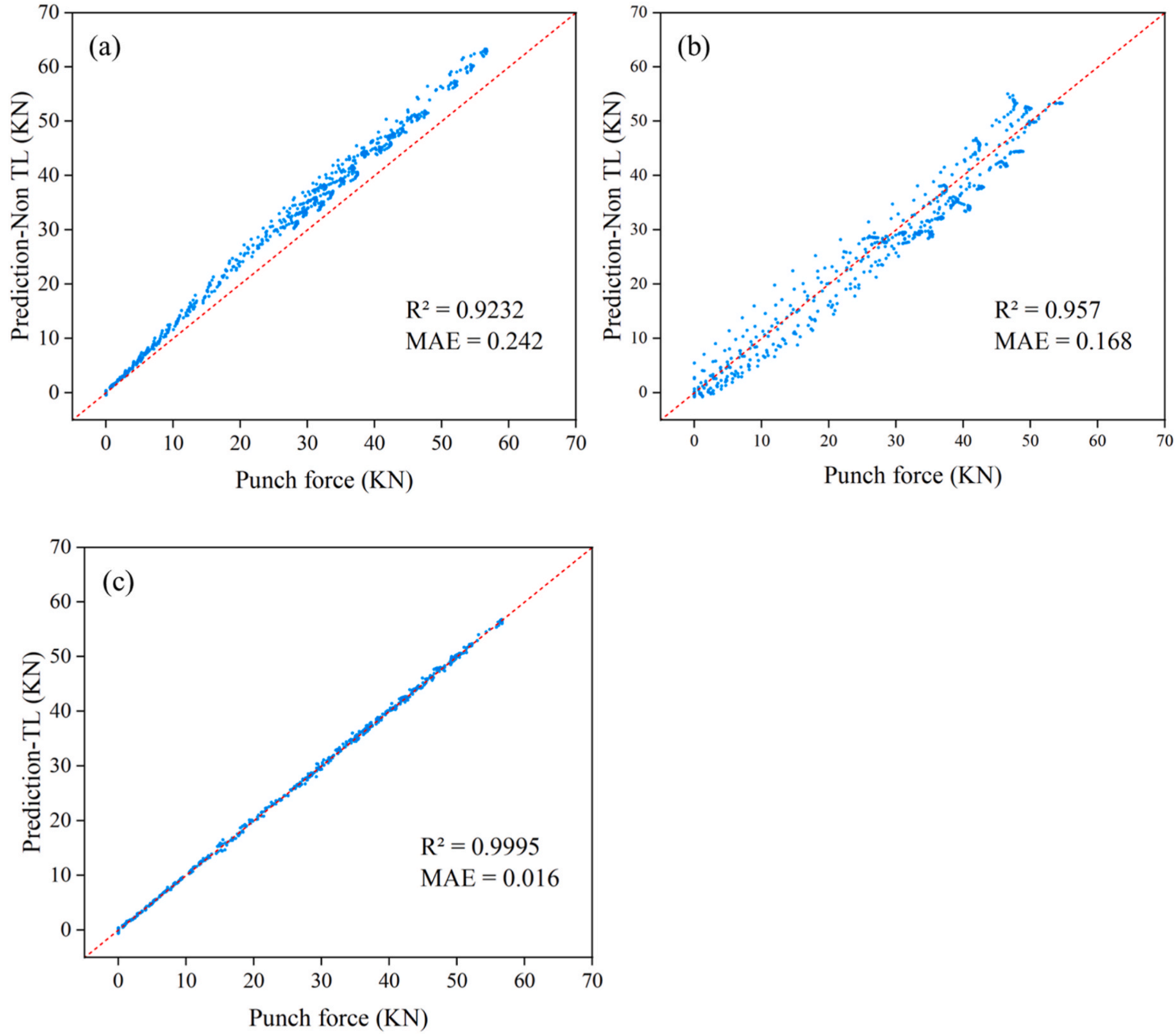

Fig. 8. Forming force predicted by: (a) analytical solution, (b) Non-TL model and (c) TL model with the same DNN architecture.

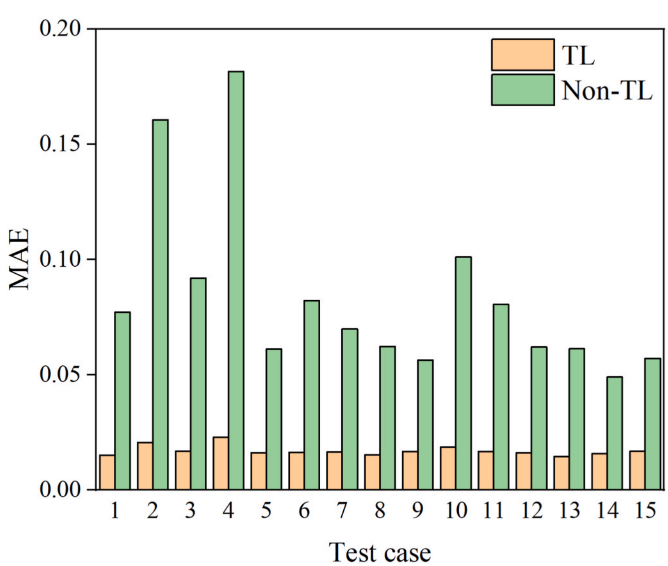

Fig. 9. MAE of TL and Non-TL models randomly partitioning the dataset 15 times.

after fine-tuning with the TL strategy, this error is effectively reduced. This shows that pre-training injects an initial representation of deep drawing physics into the network and plays a strong regularization role in the fine-tuning process.

To rigorously assess the TL model's accuracy and exclude overfitting, we conducted a 15-fold random sub-sampling experiment on our FEM dataset. In each fold, 8 curves were randomly chosen for training and the remaining 17 curves for testing. Both the TL and Non-TL models were retrained and evaluated per fold, with the Non-TL model's hyper-parameters optimized via Optuna before each run.

Fig. 9 plots the resulting MAE distributions across all trials. The TL model consistently achieves a low mean MAE of 0.0169 with a tight variance, underscoring both high accuracy and robust stability. In stark contrast, the Non-TL model's mean MAE of 0.0835 is accompanied by large fluctuations, indicating that its performance is highly sensitive to the particular training split. The Non-TL model's unstable MAE behavior reveals its overreliance on sparse training sets, leading to inconsistent predictions. The TL model's low variance demonstrates that pre-training on analytical data provides a strong inductive bias, enabling reliable performance even when target data are limited and randomly chosen.

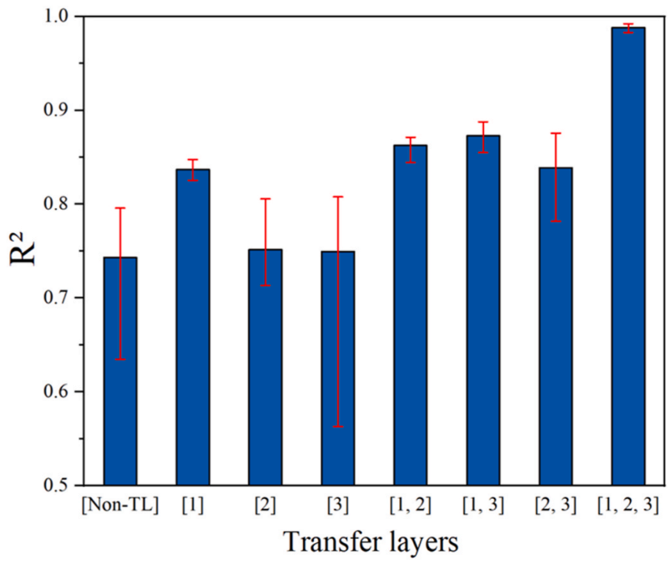

Fig. 10. R^2 when accepting different hidden layer parameters from the pre-trained model.

5.3. Performance of TL model

To dissect which pre-trained representations most benefit the target task, we investigated the effectiveness of the TL model by examining the impact of transferring different hidden layer on R^2 during the transfer process. Using the base network with three hidden layers, we conducted fine-tuning and the results are shown in Fig. 10, four variants of fine-tuning are worth noting: No layers transferred (all weights randomly initialized) average $R^2 = 0.743$; Only Layer 1 transferred average $R^2 = 0.837$; Layers 1 + 2 transferred average $R^2 = 0.863$; All three layers transferred average $R^2 = 0.988$.

The results show that transferring only the first hidden layer yields a dramatic jump from $R^2 = 0.743$ to 0.837, nearly matching the benefit of transferring two layers. This underscores that early layers capture broad and general features that are highly reusable across domains. Transferring only Layer 2 or Layer 3 fails to improve performance. This is because subsequent deeper hidden layers perform complex nonlinear transformations on the general features, gradually converting them from general to specific, and nonlinear feature combinations fine-tuned to the source task's peculiarities, offering little independent value for the new domain. Performance steadily improves as more pre-trained layers are included, peaking at $R^2 = 0.988$ when all layers are transferred. This indicates that, while deeper layers are specialized, their adaptation still contributes positively when fine-tuned collectively.

We systematically investigated how freezing different subsets of pre-trained layers affects model accuracy, using MAE as the performance metric (Fig. 11). All experiments began by transferring the full set of pre-trained parameters into the target model. Subsequently, we either froze or unfroze layers according to the following configurations.

1. All layers frozen: Only the final output layer was retrained on the FEM data, yielding the poorest predictive accuracy;
2. Progressive unfreezing: We sequentially allowed deeper layers to be fine-tuned, unfreezing layer 3 and layers 2–3;
3. Full fine-tuning: All layers were updated on the target data, producing the lowest MAE of 0.017.

These results show that keeping all hidden layers fixed forces the model to rely solely on source-task representations, which results in poor accuracy. Additional unfrozen layer consistently reduced MAE, demonstrating that successive fine-tuning unlocks progressively richer, task-relevant features. Full fine-tuning achieved the best performance, confirming that unfreezing more layers produces the most accurate and robust force predictions when target-domain features are significantly different from the source.

Fig. 12 illustrates the trend of MAE as the TL model is trained with varying numbers of FEM datasets. It can be observed that the MAE of the directly trained DNN model gradually decreases as the data volume increases, while the TL model consistently outperforms the Non-TL model under all data conditions. Specifically, when fewer than four datasets are used, the DNN model fails to converge effectively, whereas the TL model demonstrates excellent predictive performance. Notably, when only one FEM dataset is used for training, the TL model achieves an MAE as low as 0.108 and an R^2 of 0.98. As the amount of training data increases, especially when using five or more data sets, R^2 stabilizes above 0.9994, very close to 1. This demonstrates the significant advantages and superior performance of the TL method when dealing with small datasets. This result highlights the effectiveness of TL in improving model prediction accuracy. TL can leverage the knowledge learned in the pre-trained model to achieve excellent results in the target task, even with a small amount of training data. In contrast, directly trained DNN

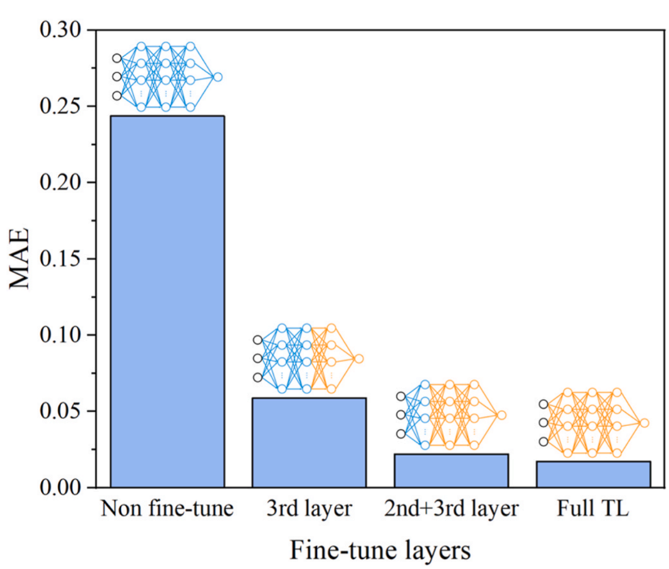

Fig. 11. MAE of different fine-tuning strategies after migrating all pre-trained parameters.

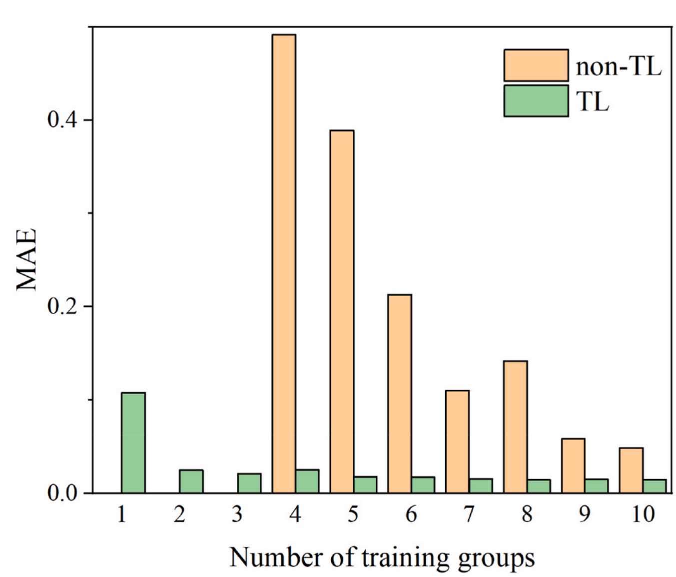

Fig. 12. MAE of TL and Non-TL with different numbers of training data sets.

Table 2
Comparison of CPU time between FEM and TL.

Method	Time
FEM	6433 s
Pre-train model	108 s
TL model	7.25 s–16.04 s

models, lacking the support of pre-trained knowledge, require large amounts of training data to achieve relatively high-performance levels. In conclusion, this study, by comparing the performance of TL models and directly trained models under different amounts of training data, further confirms the crucial role and advantages of TL in improving model prediction performance in small dataset environments. This provides strong support for the promotion and implementation of TL in practical applications, especially when data is limited. TL is undoubtedly an efficient and reliable model training strategy in such scenarios.

The above analysis confirms that the knowledge TL model has the advantage of improving prediction accuracy. This paper also considers the advantage of TL in reducing training time. Table 2 shows the CPU time required for FEM and TL. In the deep drawing process, the blank undergoes large plastic deformation involving complex loading conditions and contact problems, which often leads to slow convergence during the calculation process. It takes about 6433 s to complete a task using FEM, while the training time for the pre-trained model is only 108 s. Additionally, the training time for the network after knowledge transfer with small batches of data is even shorter, reflecting its excellent computational efficiency. The reason why the TL model significantly reduces training time is that it effectively utilizes the knowledge obtained in the pre-trained model, allowing it to quickly adjust parameters in the target task and adapt to new datasets. In contrast, FEM requires starting over with complex numerical calculations, which is time-consuming. This result not only demonstrates the significant improvement in prediction performance with TL but also highlights its advantage in computational efficiency. By reducing training time, TL can greatly enhance the practical application efficiency of the model, making it more practical and economical for industrial applications.

5.4. Knowledge transfer for non-axisymmetric deep drawing process

To evaluate the generalizability of our TL framework beyond axisymmetric cups, we next apply it to a rectangular deep-drawing task. Six FEM-generated force–stroke data are used to train both (i) a standalone DNN (with hyperparameters tuned via Optuna) and (ii) the TL model (pre-trained on analytical cup data and fine-tuned on the same six rectangular cases). The remaining 11 sets of FEM data served as an unseen test set. The results are shown in Fig. 13.

As it can be seen from Fig. 13 (a), although the optimized DNN captures the overall upward trend of the punch force ("*Fp*") versus stroke ("*h*"), its predictions exhibit noticeable scatter around the true curves, yielding an R^2 of only 0.9236. By contrast, the TL model produces force–stroke data that almost perfectly overlap the FEM results, achieving an R^2 of 0.9985. These results demonstrate that features learned from axisymmetric deep drawing can be effectively transferred to a distinct rectangular geometry, even with only six fine-tuning samples. The TL strategy thus provides a powerful means to extend high-accuracy force prediction to new part shapes with minimal additional data.

6. Discussion

In the field of machine learning, transfer learning has the advantages of improving prediction accuracy and shortening training time. Sheet metal deep drawing is a highly nonlinear problem. Analytical equations have the advantage of fast computational efficiency, but they are usually calculated based on assumptions and simplifications. Such results usually have large deviations or are not applicable to certain problems. FEM can simulate deep drawing behavior more accurately, but it often requires large computing resources. Transfer learning provides a bridge between the two methods, connecting the two knowledge models in the form of data. We can obtain a large amount of forming force data from mature analytical methods for pre-training basic neural networks. Through these data, we can fully learn the general characteristic relationship of forming force-displacement, so we can reduce the demand for high-fidelity data and only need less FEM data to adapt effectively. And if new data, such as experimental data, is obtained, it can be used to fine-tune the saved model without having to create the model from scratch, so that it can acquire knowledge from the data of different methods. This can make full use of existing data in the industrial field, and of course it is necessary to identify whether there is a certain correlation.

The principal novelty of our method is its purely data-driven use of analytical formulations: unlike Physics-Informed Neural Networks (PINNs) that integrate hard physical constraints into the architecture or loss, our framework treats analytical data simply as a rich source domain. This decoupling yields a flexible, modular pipeline that can be easily updated as new FEM or experimental data become available, without retraining from scratch.

Our layer freezing experiments reveal that unfreezing more layers invariably enhances predictive accuracy, with full fine-tuning achieving

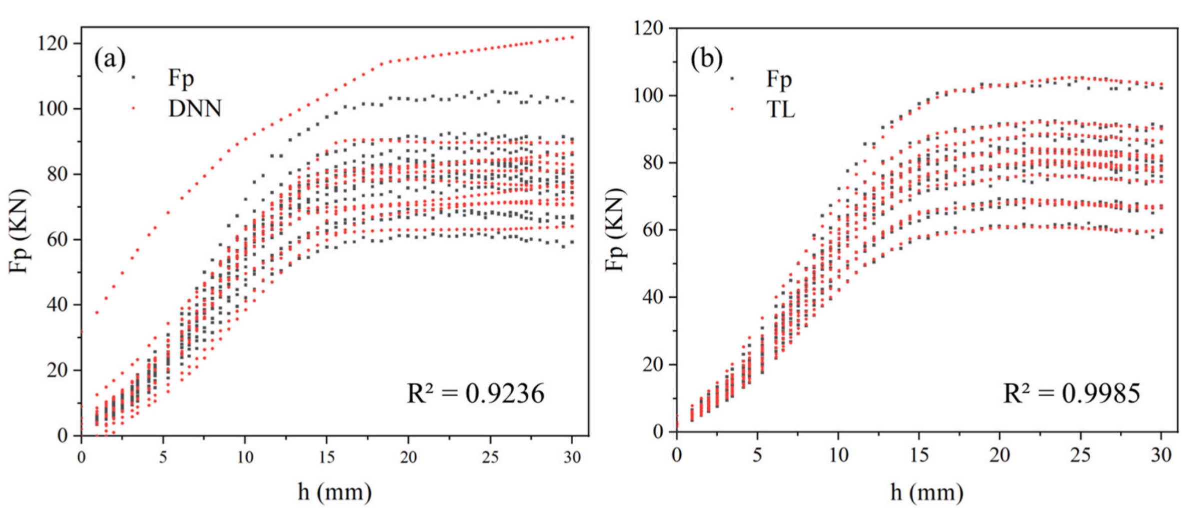

Fig. 13. Forming force predicted by (a) DNN and (b) TL.

the lowest MAE and highest R^2. Yet, freezing the first hidden layer alone preserves excellent performance, underscoring that early layers capture broadly transferable, physics-driven features. This suggests a pragmatic compromise: one may freeze foundational layers to accelerate training while retaining most of the TL benefit. However, the data volume in this study is not large, so we mainly focus on improving the prediction accuracy. Therefore, full model fine-tuning is used as the main migration strategy.

Despite its strengths, this approach carries the following risk: the analytical pre-training may embed biases that propagate through TL if its simplifying assumptions (axisymmetric, constant thickness, uniform friction) deviate significantly from the target application. Future work must therefore assess TL's robustness across alternative analytical models and explore mitigation strategies to curtail inherited errors. In addition, our analytical pre-training for axisymmetric cups and its extension to deep drawing of rectangular plates can effectively predict the forming forces of new shapes, proving that the force responses of the two are very similar. However, for more complex parts, further exploration of the correlation is needed. Regarding the model architecture, sequence-based models, i.e. Recurrent Neural Network/Gated Recurrent Unit (RNN/GRU), may be more effective in capturing the temporal correlations on the force-displacement curves than pointwise DNNs and deserve further investigation. Also, the characterization scope of this paper is limited, as only the coefficient of friction and thickness are considered. Other process variables (e.g., strain rate, anisotropy, temperature) should be added to the TL framework and tested for scalability.

7. Conclusions

This study proposes a TL framework for metal deep drawing, aiming to achieve high prediction accuracy of forming force under conditions of data scarcity. This framework combines analytically generated deep-drawing force data with a minimal set of high-fidelity FEM data to predict forming force under severe data constraints. By pre-training a DNN on energy-based axisymmetric forming equations and fine-tuning on only eight FEM curves, our TL approach achieved near-perfect accuracy ($R^2 = 0.9995$ and MAE = 0.016), significantly better than both the analytical baseline ($R^2 = 0.9232$, MAE = 0.242) and a directly trained DNN ($R^2 = 0.957$, MAE = 0.168). This is mainly attributed to the inductive bias imparted by analytical pre-training, which equips the network with generalized physical features that facilitate rapid convergence and robust generalization. Our layer-wise analyses further reveal that the first hidden layer captures the bulk of these transferable features (R^2 jump from 0.743 to 0.837), while subsequent layers refine task-specific nuances. This suggests a principled, data-efficient strategy: freeze foundational layers to preserve broad physics knowledge, and fine-tune deeper layers to adapt to new domains. Beyond accuracy, TL also significantly reduces computational costs: compared to a 6,433 s FEM run, pre-training and TL require only 108 s and 7–16 s, respectively. By extending our framework to rectangular cups, we demonstrated that features learned from axisymmetric drawing can transfer effectively to non-axisymmetric tasks, and the TL model ($R^2 = 0.9985$) can effectively eliminate the scatter observed in directly trained DNNs under data scarcity ($R^2 = 0.9236$), improving accuracy. In conclusion, this work establishes TL as a powerful alternative modeling paradigm for metal forming. It leverages existing analytical and numerical knowledge to deliver high-fidelity predictions with minimal additional data, offering a compelling route to accelerate die design and process optimization under data-scarce conditions.

CRediT authorship contribution statement

Yingjian Guo: Writing – original draft, Methodology, Conceptualization. **Gregor Kosec:** Writing – review & editing. **Lihua Wang:** Writing – review & editing, Supervision. **Magd Abdel Wahab:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The authors would like to acknowledge the financial support of the grants from the China Scholarship Council (CSC 202206020076). This work is supported by the National Natural Science Foundation of China (project no. 12272270), the Shanghai Pilot Program for Basic Research and the Fundamental Research Funds for the Central Universities.

Data availability

Data will be made available on request.

References

Assempour, A., Taghipour, E., 2011. The effect of normal stress on hydro-mechanical deep drawing process. Int. J. Mech. Sci. 53 (6), 407–416. https://doi.org/10.1016/j.ijmecsci.2011.03.003.

Atrian, A., Fereshteh-Saniee, F., 2013. Deep drawing process of steel/brass laminated sheets. Compos. B Eng. 47, 75–81. https://doi.org/10.1016/j.compositesb.2012.10.023.

Babu, K.V., Narayanan, R.G., Kumar, G.S., 2010. An expert system for predicting the deep drawing behavior of tailor welded blanks. Expert Syst. Appl. 37 (12), 7802–7812. https://doi.org/10.1016/j.eswa.2010.04.059.

Basak, S., Panda, S.K., 2023. Use of uncoupled ductile damage models for fracture forming limit prediction during two-stage forming of aluminum sheet material. J. Manuf. Process. 97, 185–199. https://doi.org/10.1016/j.jmapro.2023.04.042.

Basak, S., Panda, S.K., Lee, M.G., 2020. Formability and fracture in deep drawing sheet metals: extended studies for pre-strained anisotropic thin sheets. Int. J. Mech. Sci. 170, 105346. https://doi.org/10.1016/j.ijmecsci.2019.105346.

Bonatti, C., Mohr, D., 2021a. One for all: universal material model based on minimal state-space neural networks. Sci. Adv. 7 (26), 3658. https://doi.org/10.1126/sciadv.abf3658.

Bonatti, C., Mohr, D., 2021b. Neural network model predicting forming limits for Bi-linear strain paths. Int. J. Plast. 137, 102886. https://doi.org/10.1016/j.ijplas.2020.102886.

Cheng, J., De Waele, W., 2024. Weighted average algorithm: a novel meta-heuristic optimization algorithm based on the weighted average position concept. Knowl. Base Syst. 305, 112564. https://doi.org/10.1016/j.knosys.2024.112564.

Choi, T.H., Huh, H., 1999. Sheet metal forming analysis of planar anisotropic materials by a modified membrane finite element method with bending effect. J. Mater. Process. Technol. 89–90, 58–64. https://doi.org/10.1016/S0924-0136(99)00050-3.

Choi, H., Koç, M., Ni, J., 2007. A study on the analytical modeling for warm hydro-mechanical deep drawing of lightweight materials. Int. J. Mach. Tool Manufact. 47 (11), 1752–1766. https://doi.org/10.1016/j.ijmachtools.2006.12.005.

Cwiekala, T., Brosius, A., Tekkaya, A.E., 2011. Accurate deep drawing simulation by combining analytical approaches. Int. J. Mech. Sci. 53 (5), 374–386. https://doi.org/10.1016/j.ijmecsci.2011.02.007.

Daxin, E., Mizuno, T., Li, Z.G., 2008. Stress analysis of rectangular cup drawing. J. Mater. Process. Technol. 205 (1), 469–476. https://doi.org/10.1016/j.jmatprotec.2007.11.274.

Dong, F., Huang, S.Q., Yi, Y.P., He, H.L., Huang, K., Wang, C.G., Gao, S.L., Jia, Y.Z., Yu, W.W., 2023. Enhanced deep drawing formability and deformation mechanism of aluminum alloy at cryogenic temperature. J. Alloys Compd. 968, 171992. https://doi.org/10.1016/j.jallcom.2023.171992.

Du, B., Song, P.F., Li, H., Dong, G.J., Cao, M.Y., Zhao, C.C., 2021. Establishment of sheet metal forming wrinkling limit diagram (WLD) and research on the consistency of WLDs in different processes. Thin-Walled Struct. 164, 107770. https://doi.org/10.1016/j.tws.2021.107770.

Duan, S.L., Kozjek, D., Mehr, E., Anders, M., Cao, J., 2024. Forming force prediction in double-sided incremental forming via GNN-based transfer learning. J. Manuf. Process. 120, 867–877. https://doi.org/10.1016/j.jmapro.2024.04.093.

El-Brawany, M.A., Ibrahim, D.A., Elminir, H.K., Elattar, H.M., Ramadan, E.A., 2023. Artificial intelligence-based data-driven prognostics in industry: a survey. Comput. Ind. Eng. 184, 109605. https://doi.org/10.1016/j.cie.2023.109605.

Fazli, A., Arezoo, B., 2012. Prediction of limiting drawing ratio considering the effective parameters of die arc region. J. Mater. Process. Technol. 212 (4), 745–751. https://doi.org/10.1016/j.jmatprotec.2011.10.025.

Fereshteh-Saniee, F., Montazeran, M.H., 2003. A comparative estimation of the forming load in the deep drawing process. J. Mater. Process. Technol. 140 (1), 555–561. https://doi.org/10.1016/S0924-0136(03)00793-3.

Firouzjaei, M.K., Naeini, H.M., Kasaei, M.M., Mirnia, M.J., da Silva, L.F.M., 2024. Microscale modeling of the ductile fracture behavior of thin stainless steel sheets. Thin-Walled Struct. 196, 111457. https://doi.org/10.1016/j.tws.2023.111457.

Gao, M.D., Huang, H.H., Wang, Q.Y., Liu, Z.F., Li, X.Y., 2018. Energy consumption analysis on sheet metal forming: focusing on the deep drawing processes. Int. J. Adv. Des. Manuf. Technol. 96, 3893–3907. https://doi.org/10.1007/s00170-018-1735-1.

Giorleo, L., Ceretti, E., 2022. Deep drawing punches produced using fused filament fabrication technology: performance evaluation. J. Manuf. Process. 84, 1–9. https://doi.org/10.1016/j.jmapro.2022.09.054.

Guo, Y.G., Wang, C., Han, S.T., Kosec, G., Zhou, Y.L., Wang, L.H., Wahab, M.A., 2025. A deep neural network model for parameter identification in deep drawing metal forming process. J. Manuf. Process. 133, 380–394. https://doi.org/10.1016/j.jmapro.2024.11.067.

Haghighat, E., Abouali, S., Vaziri, R., 2023. Constitutive model characterization and discovery using physics-informed deep learning. Eng. Appl. Artif. Intell. 120, 105828. https://doi.org/10.1016/j.engappai.2023.105828.

Irthiea, I., Green, G., Hashim, S., Kriama, A., 2014. Experimental and numerical investigation on micro deep drawing process of stainless steel 304 foil using flexible tools. Int. J. Mach. Tool Manufact. 76, 21–33. https://doi.org/10.1016/j.ijmachtools.2013.09.006.

Kardan, M., Parvizi, A., Askari, A., 2018. Experimental and finite element results for optimization of punch force and thickness distribution in deep drawing process. Arabian J. Sci. Eng. 43 (3), 1165–1175. https://doi.org/10.1007/s13369-017-2783-9.

Kim, H., Sung, J.H., Sivakumar, R., Altan, T., 2007. Evaluation of stamping lubricants using the deep drawing test. Int. J. Mach. Tool Manufact. 47 (14), 2120–2132. https://doi.org/10.1016/j.ijmachtools.2007.04.014.

Kim, H.S., Koç, M., Ni, J., 2008. Development of an analytical model for warm deep drawing of aluminum alloys. J. Mater. Process. Technol. 197 (1), 393–407. https://doi.org/10.1016/j.jmatprotec.2007.06.046.

Lee, S., Lim, Y., Galdos, L., Lee, T., Quagliato, L., 2024. Gaussian process regression-driven deep drawing blank design method. Int. J. Mech. Sci. 265, 108898. https://doi.org/10.1016/j.ijmecsci.2023.108898.

Leu, D.K., 1997. Prediction of the limiting drawing ratio and the maximum drawing load in cup-drawing. Int. J. Mach. Tool Manufact. 37 (2), 201–213. https://doi.org/10.1016/0890-6955(95)00107-7.

Leu, D.K., 1999. The limiting drawing ratio for plastic instability of the cup-drawing process. J. Mater. Process. Technol. 86 (1), 168–176. https://doi.org/10.1016/S0924-0136(98)00307-0.

Li, W., Bazant, M.Z., Zhu, J., 2021. A physics-guided neural network framework for elastic plates: comparison of governing equations-based and energy-based approaches. Comput. Methods Appl. Mech. Eng. 383, 113933. https://doi.org/10.1016/j.cma.2021.113933.

Link, P., Penter, L., Rückert, U., Klingel, L., Verl, A., Ihlenfeldt, S., 2025. Real-time quality prediction and local adjustment of friction with digital twin in sheet metal forming. Robot. Comput. Integrated Manuf. 91, 102848. https://doi.org/10.1016/j.rcim.2024.102848.

Liu, Z., Jiang, M., Luo, T., 2020. Leverage electron properties to predict phonon properties via transfer learning for semiconductors. Sci. Adv. 6 (45), eabd1356. https://doi.org/10.1126/sciadv.abd1356.

Liu, Z.Y., Xiong, B.Q., Li, X.W., Yan, L.Z., Li, Z.H., Zhang, Y.A., Liu, H.W., 2019. Deep drawing of 6A16 aluminum alloy for automobile body with various blank-holder forces. Rare Met. 38 (10), 946–953. https://doi.org/10.1007/s12598-018-1146-3.

Luyen, T.T., Tong, V.C., Nguyen, D.T., 2022. A simulation and experimental study on the deep drawing process of SPCC sheet using the graphical method. Alex. Eng. J. 61 (3), 2472–2483. https://doi.org/10.1016/j.aej.2021.07.009.

Medell-Castillo, H.I., García-Zugasti, P.D., de Lange, D.F., Colorado-Alonso, F.J., 2013. Analysis of the allowable deep drawing height of rectangular steel parts. Int. J. Adv. Des. Manuf. Technol. 66 (1), 371–380. https://doi.org/10.1007/s00170-012-4331-9.

Modanloo, V., Jang, S., Lee, T., Quagliato, L., 2025. Gradient enhanced-expert informed neural network (GE-EINN) for forming depth prediction from a small-scale metal stamping dataset. J. Manuf. Process. 140, 224–240. https://doi.org/10.1016/j.jmapro.2025.02.052.

Momeni, A., Rahmani, B., Malléjac, M., del Hougne, P., Fleury, R., 2023. Backpropagation-free training of deep physical neural networks. Science 382 (6676), 1297–1303. https://doi.org/10.1126/science.adi8474.

Müller, M., Uhlmann, L., Herrig, T., Bergs, T., 2024. Numerical analysis of the dependence of damage on friction during deep drawing of asymmetric geometries. Advances in Industrial and Manufacturing Engineering 8, 100136. https://doi.org/10.1016/j.aime.2024.100136.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Qu, T.M., Zhao, J.D., Guan, S.H., Feng, Y.T., 2023. Data-driven multiscale modelling of granular materials via knowledge transfer and sharing. Int. J. Plast. 171, 103786. https://doi.org/10.1016/j.ijplas.2023.103786.

Rivas-Menchi, A., Medellín-Castillo, H.I., de Lange, D.F., García-Zugasti, P.D.J., 2018. Performance evaluation of analytical expressions for cylindrical and rectangular deep drawing force estimation. J. Manuf. Process. 36, 340–350. https://doi.org/10.1016/j.jmapro.2018.10.033.

Sabet, A.S., Domitner, J., Öksüz, K.I., Hodžić, E., Torres, H., Ripoll, M.R., Sommitsch, C., 2021. Tribological investigations on aluminum alloys at different contact conditions for simulation of deep drawing processes. J. Manuf. Process. 68, 546–557. https://doi.org/10.1016/j.jmapro.2021.05.050.

Samad, A., Thakur, A.K., Basak, S., 2025. Formability prediction of anisotropic thin sheets using machine learning framework: influence of pre-strain and temperature. J. Mater. Eng. Perform. https://doi.org/10.1007/s11665-025-11093-w.

Shutov, A.V., Silbermann, C.B., Ihlemann, J., 2015. Ductile damage model for metal forming simulations including refined description of void nucleation. Int. J. Plast. 71, 195–217. https://doi.org/10.1016/j.ijplas.2015.03.003.

Storm, J., Rocha, I.B.C.M., van der Meer, F.P., 2024. A microstructure-based graph neural network for accelerating multiscale simulations. Comput. Methods Appl. Mech. Eng. 427, 117001. https://doi.org/10.1016/j.cma.2024.117001.

Susmel, L., 2024. Estimating notch fatigue limits via a machine learning-based approach structured according to the classic Kf formulas. Int. J. Fatig. 179, 108029. https://doi.org/10.1016/j.ijfatigue.2023.108029.

Vollertsen, F., Hu, Z., Niehoff, H.S., Theiler, C., 2004. State of the art in micro forming and investigations into micro deep drawing. J. Mater. Process. Technol. 151 (1), 70–79. https://doi.org/10.1016/j.jmatprotec.2004.04.266.

Wan, M., Yang, Y.Y., Li, S.B., 2001. Determination of fracture criteria during the deep drawing of conical cups. J. Mater. Process. Technol. 114 (2), 109–113. https://doi.org/10.1016/S0924-0136(01)00730-0.

Wang, C.G., Ma, R., Zhao, J., Zhao, J., 2017. Calculation method and experimental study of coulomb friction coefficient in sheet metal forming. J. Manuf. Process. 27, 126–137. https://doi.org/10.1016/j.jmapro.2017.02.016.

Weng, H.B., Yuan, H., 2023. Quantitative representation of directional microstructures of single-crystal superalloys in cyclic crystal plasticity based on neural networks. Int. J. Plast. 170, 103757. https://doi.org/10.1016/j.ijplas.2023.103757.

Wu, L.J., 2022. Research on friction mechanism of deep drawing for cylindrical parts and hydro-pulsating lubrication technology. PhD Thesis. Yanshan University, China. https://d.wanfangdata.com.cn/thesis/D02948360.

Yang, Z., Yu, C.H., Buehler, M.J., 2021. Deep learning model to predict complex stress and strain fields in hierarchical composites. Sci. Adv. 7 (15), 7416. https://doi.org/10.1126/sciadv.abd7416.

Yoshihara, S., Manabe, K., Nishimura, H., 2005. Effect of blank holder force control in deep-drawing process of magnesium alloy sheet. J. Mater. Process. Technol. 170 (3), 579–585. https://doi.org/10.1016/j.jmatprotec.2005.06.028.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How Transferable Are Features in Deep Neural Networks?, vol. 27.

Zhao, J., Qin, S.J., Cao, H.Q., Li, S.B., 1998. Analytic description of the intelligent deep drawing process for axisymetric-curved workpieces 5 (4), 47–58.

Zhao, J., Wang, F., 2005. Parameter identification by neural network for intelligent deep drawing of axisymmetric workpieces. J. Mater. Process. Technol. 166 (3), 387–391. https://doi.org/10.1016/j.jmatprotec.2004.08.020.

Zhao, J., Cao, H.Q., Ma, L.X., Wang, F.Q., Li, S.B., 2004. Study on intelligent control technology for the deep drawing of an axi-symmetric shell part. J. Mater. Process. Technol. 151 (1), 98–104. https://doi.org/10.1016/j.jmatprotec.2004.04.023.

Zhao, H., Ma, X.G., Wang, Z.H., Jiang, Z.Y., Zhou, C.L., Zhao, J.W., 2024. Surface roughness evolution and heterogeneous plastic deformation of austenitic stainless steel during micro deep drawing: modeling and experiment. Int. J. Plast. 176, 103964. https://doi.org/10.1016/j.ijplas.2024.103964.

Zheng, K.L., Lee, J.Y., Lin, J.G., Dean, T.A., 2017. A buckling model for flange wrinkling in hot deep drawing aluminium alloys with macro-textured tool surfaces. Int. J. Mach. Tool Manufact. 114, 21–34. https://doi.org/10.1016/j.ijmachtools.2016.12.008.